BIOMETRICS

VOLUME 11: 1955

Numbers 1–4

MULTIPLE RANGE AND MULTIPLE F TESTS*

DAVID B. DUNCAN**

Virginia Polytechnic Institute
Blacksburg, Virginia

1. INTRODUCTION

The common practice for testing the homogeneity of a set of n treatment means in an analysis of variance is to use an F (or z) test. This procedure has special desirable properties for testing the homogeneity hypothesis that the n population means concerned are equal. An F test alone, however, generally falls short of satisfying all of the practical requirements involved. When it rejects the homogeneity hypothesis, it gives no decisions as to which of the differences among the treatment means may be considered significant and which may not.

To illustrate, Table I shows results of a barley grain yield experiment conducted by E. Shulkcum of this Institute at Accomac, Virginia, in 1951. Seven varieties, A, B, $\cdots$, G, were replicated six times in a randomized block design. The F ratio (in section b) for testing the homogeneity of the varietal means is highly significant. This indicates that one or more of the differences among the means are significant but it does not specify which ones.

TABLE I.   BARLEY GRAIN YIELDS IN BUSHELS PER ACRE

a) *Varietal Means Ranked in Order*

A	F	G	D	C	B	E
49.6	58.1	61.0	61.5	67.6	71.2	71.3

b) *Analysis of Variance*

Source	d.f.	m.s.	F
Between varieties	6	366.97	4.61**
Between blocks	5	141.95	
Error	30	79.64	

c) *Standard Error of a Varietal Mean*

$$s_m = \sqrt{79.64/6} = 3.643 \qquad (n_2 = 30)$$

The problem we wish to consider is that of testing these differences more specifically. Several test procedures have been proposed for

answering this problem. The simplest of these is one which is often termed the *least-significant-difference* (or *L.S.D.*) *test*. This has developed from a brief discussion of the problem by R. A. Fisher (9, section 24) and is described in detail by several authors, for example, Paterson (14, pp. 38-42) and Davies (4, section 5.28). In this test, the difference between any two means is declared significant, at the 5% level, say, if it exceeds a so-called *least significant difference* $\sqrt{2}\ ts_m$ (t being the 5% level significant value from the t distribution), and provided also that the F test for the homogeneity of the n means involved is significant. If the F test is not significant, none of the differences is significant irrespective of its magnitude relative to the least significant difference.

Many other tests have also been proposed for solving this problem, including several put forward within the last year or two. Further tests are being developed at the present time. Originators of these, not to mention all, include D. T. Sawkins (18), D. Newman (12), D. B. Duncan (5-8), J. W. Tukey (21-23), H. Scheffé (19), M. Keuls (10), S. N. Roy, R. C. Bose (17), H. O. Hartley (25), and J. Cornfield, M. Halperin, S. Greenhouse (3). Unfortunately, these tests vary considerably and it is difficult for the user to decide which one to choose for any given problem.

One objective of this paper is to consider several of the procedures which have been proposed and to illustrate their basic points of difference, using a geometric method with simple cases involving only three means. A second objective is to present certain simple extensions of the concepts of power and significance which are useful in analyzing these procedures. The development of the simple case examples and the latter general concepts will point the way to a clearer evaluation of the relative properties and merits of the procedures in general and should help the user in making a choice among the available procedures. The final objective is to present a new multiple range test (8) which combines the features considered to be the best from the previously proposed tests.

2. THE NEW MULTIPLE RANGE TEST

Before discussing the general problem in more detail, it may be helpful to look ahead at an example of the application of one of the tests. An example of the proposed new test will be used for this purpose. This *new multiple range test*, as it will be termed, combines the simplicity and speed of application of a test proposed by Newman (12) and Keuls (10) with most of the power advantages of the multiple comparisons test previously proposed by the author (6, 7). For the example, we shall consider the application of a 5% level test to the varietal yield means in Table I.

TABLE II. SIGNIFICANT STUDENTIZED RANGES FOR A 5% LEVEL NEW* MULTIPLE RANGE TEST

n_2 \\ p	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.44	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.47	3.47	3.47	3.47
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.47	3.47	3.47	3.47
24	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.47	3.47	3.47	3.47
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47	3.47	3.47
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47	3.47	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47	3.53	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47	3.61	3.67

*Using special protection levels based on degrees of freedom.

TABLE III. SIGNIFICANT STUDENTIZED RANGES FOR A 1% LEVEL NEW* MULTIPLE RANGE TEST

p / n_2	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3	9.3	9.3
4	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5	7.5	7.5
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8	6.8	6.8
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.1	6.2	6.2	6.3	6.3	6.3	6.3
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0	6.0	6.0
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8	5.8	5.8
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7	5.7	5.7
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55	5.55	5.55
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39	5.39	5.39
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26	5.26	5.26
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15	5.15	5.15
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07	5.07	5.07
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00	5.00	5.00
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94	4.94	4.94
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89	4.89	4.89
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85	4.85	4.85
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82	4.82	4.82
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79	4.79	4.79
22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75	4.75	4.75
24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72	4.74	4.74
26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69	4.73	4.73
28	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67	4.72	4.72
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65	4.71	4.71
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59	4.69	4.69
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53	4.64	4.66
100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48	4.64	4.65
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41	4.60	4.68

*Using special protection levels based on degrees of freedom.

The data necessary to perform the test are: (a) the means as shown in Table I; (b) the standard error of each mean, $s_m = 3.643$ and (c) the degrees of freedom on which this standard error is based, $n_2 = 30$.

First, a table (Table II) of special *significant studentized ranges* for a 5% level test is entered at the row for $n_2 = 30$ degrees of freedom, and significant studentized ranges are extracted for samples of sizes $p = 2, 3, 4, 5, 6$ and 7. The values obtained in this way are 2.89, 3.04, 3.12, 3.20, 3.25 and 3.29 respectively. (Table III shows the significant studentized ranges which would be used for a 1% level test.)

The significant studentized ranges are then each multiplied by the standard error, $s_m = 3.643$, to form what may be called *shortest significant ranges*. The shortest significant ranges R_2, R_3, $\cdots$, R_7 are recorded at the top of a worksheet as shown in Table IV.

As a final preparatory step it is convenient to display the means in ranked order from left to right, spaced so that the distances between them are very roughly proportional to their numerical differences. This may be done on the worksheet immediately under the shortest significant ranges as in Table IV. The lines underscoring the means indicate the results and are added as the test proceeds.

TABLE IV. WORKSHEET

a) *Shortest Significant Ranges*

p:	(2)	(3)	(4)	(5)	(6)	(7)
R_p :	10.53	11.07	11.37	11.66	11.84	11.99

b) *Results*

Varieties:	A	F	G	D	C	B	E
Means:	49.6	58.1	61.0	61.5	67.6	71.2	71.3

Note: Any two means *not underscored* by the same line are *significantly different.*

Any two means *underscored* by the same line are *not significantly different.*

We now set out to test the differences in the following order: the largest minus the smallest, the largest minus the second smallest, up to the largest minus the second largest; then the second largest minus the smallest, the second largest minus the second smallest, and so on, finishing with the second smallest minus the smallest. Thus, in the case of this example the order for testing is: $E - A, E - F, E - G, E - D, E - C, E - B; B - A, B - F, B - G, B - D, B - C; C - A, C - F, C - G, C - D; D - A, D - F, D - G; G - A, G - F$; and finally $F - A$.

With only one exception, given below, *each difference is significant if it exceeds the corresponding shortest significant range; otherwise it is not significant.* Because $E - A$ is the range of seven means, it must exceed $R_7 = 11.99$, the shortest significant range of seven means, to be significant; because $E - F$ is the range of six means, it must exceed $R_6 = 11.84$, the shortest significant range for six means, to be significant; and so on. *Exception*: The sole exception to this rule is that *no difference between two means can be declared significant if the two means concerned are both contained in a subset* of the means which has a non-significant range.*

Because of this exception, as soon as a non-significant difference is found between two means, it is convenient to group these two means and all of the intervening means together by underscoring them with a line, as shown for the means $\{G, D, C, B, E\}$, for example, in Table IV. The remaining differences between all members of a subset underscored in this way are not significant according to the exception rule. Thus they need not, and should not, be tested against shortest significant ranges.

The details of the test are as follows:

1) $E - A = 21.7 > 11.99$; thus $E - A$ is significant.

2) $E - F = 13.2 > 11.84$; thus $E - F$ is significant.

3) $E - G = 10.3 < 11.66$; thus $E - G$ is not significant, and hence $E - D, E - C, E - B; B - G, B - D, B - C; C - G, C - D$; and $D - G$ are not significant by the exception rule. These results are all denoted by drawing the line under the subset $\{G, D, C, B, E\}$.

4) $B - A = 21.6 > 11.84$; thus $B - A$ is significant.

5) $B - F = 13.1 > 11.66$; thus $B - F$ is significant.

6) $B - G, B - D, B - C; C - G, C - D$; and $D - G$ are not significant from step 3. No line need be added to show this because of the line under $\{G, D, C, B, E\}$ already.

7) $C - A = 18.0 > 11.66$; thus $C - A$ is significant.

8) $C - F - 9.5 < 11.37$; thus $C - F$ is not significant; and $C - G, C - D; D - F, D - G$; and $G - F$ are not significant by the exception rule. These results are all denoted by drawing the line under the subset $\{F, G, D, C\}$.

9) $D - A = 11.9 > 11.37$; thus $D - A$ is significant.

10) $D - F$ is not significant from step 8 and $D - G$ is not significant from step 3 or 8.

11) $G - A = 11.4 > 11.07$; thus $G - A$ is significant.

12) $G - F$ is not significant from step 8.

---

*The term *subset* will be used to include the complete set where necessary, as is the case here.

13) $F - A = 8.5 < 10.53$; thus $F - A$ is not significant. The result is denoted by drawing the line under $\{A, F\}$.

Each of the steps can be done almost by inspection and the complete test takes very little time. All that is necessary for a complete recording of the result is the array of means with the lines underneath, together with the brief statement giving their interpretation, as shown in section b of Table IV.

In practice there is a short cut which can be used repeatedly to good advantage, especially when the number of means is large. Instead of starting by finding the difference $E - A$, subtract the shortest significant range for seven means from the top mean E. This gives $71.3 - 11.99 = 59.31$. Since A and F are each less than 59.31, it follows that $E - A$ and $E - F$ are both significant. This is so because the shortest significant ranges R_p become smaller with decreases in the subset size p. This takes care of steps 1 and 2 in one operation. The same idea can be used repeatedly throughout the complete application and may often eliminate many steps at a time especially in a case with a large number of means.

The foregoing provides a brief introduction to many of the features of the problem involved as well as an illustration of the proposed new multiple range test. We now begin afresh considering matters in more detail.

3. GENERAL ASSUMPTIONS AND DECISIONS

In the general problem we are given a sample of observed means, $m_1, m_2, \cdots, m_n$, which are assumed to have been drawn independently from n normal populations with "true" means, $\mu_1, \mu_2, \cdots, \mu_n$, respectively, and a common standard error σ_m. This standard error is unknown, but there is available the usual estimate s_m, which is independent of the observed means and is based on a number of degrees of freedom, denoted by n_2. (More precisely, s_m has the property that $n_2 s_m^2 / \sigma_m^2$ is distributed as χ^2 with n_2 degrees of freedom, independently of $m_1, m_2, \cdots, m_n$.)

In the simplest case, with only two means m_1 and m_2, there are three possible decisions. These are:

1) *m_1 is significantly less than m_2;*
2) *m_1 and m_2 are not significantly different;*
3) *m_2 is significantly less than m_1.*

It is convenient to denote these decisions by (1, 2), (1, 2), and (2, 1), respectively. The order of the numbers in each pair of parentheses indicates the ranking of the means except when underscored, in which case the means are not ranked.

In passing it should be noted that we do not intend to restrict consideration, as some writers have done, for example R. E. Bechhofer (1), to problems in which the middle decision $(1, 2)$ is eliminated and the investigator is obliged to make one of the two positive decisions (1, 2) or (2, 1). Problems of this type and their extensions to cases involving more than two means may be regarded as special cases of the problems treated here in which the significance level is fixed at 100% instead of the usual 5% or 1% level.

In the case $n = 3$, with three means, m_1 , m_2 , and m_3 , there are 19 possible decisions. These comprise:

a) Six decisions of the form: "m_1 *is significantly less than* m_2 , m_2 *is significantly less than* m_3 , *and* m_1 *is significantly less than* m_3 ." This joint decision may be conveniently denoted by (1, 2, 3). The remaining five denoted in the same way are (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), and (3, 2, 1).

b) Three decisions of the form: "m_1 *is significantly less than* m_2 *and* m_3 , *but* m_2 *and* m_3 *are not significantly different from one another.*" This joint decision may be denoted by $(1, \underline{2, 3})$. The remaining two denoted in the same way are $(2, \underline{1, 3})$ and $(3, \underline{1, 2})$.

c) Three decisions of the form: "m_1 *and* m_2 *are significantly less than* m_3 , *but* m_1 *and* m_2 *are not significantly different from one another.*" This one may be denoted by $(\underline{1, 2}, 3)$ and the remaining two in a similar way by $(\underline{1, 3}, 2)$ and $(\underline{2, 3}, 1)$.

d) Six decisions of the form: "m_1 *is significantly less than* m_3 , *but* m_1 *and* m_2 *are not significantly different from one another, and* m_2 *and* m_3 *are not significantly different from one another.*" This decision may be denoted by $(\underline{1, 2, 3})$ and the remainder by $(\underline{1, 3, 2})$, $(\underline{2, 1, 3})$, $(\underline{2, 3, 1})$, $(\underline{3, 1, 2})$, and $(\underline{3, 2, 1})$.

e) One decision stating: "m_1 , m_2 , *and* m_3 *are not significantly different from one another,*" which may be denoted by $(\underline{1, 2, 3})$.

The number of decisions increases very rapidly as n increases. In the general case with n means there are $n!$ decisions of the form $(1, 2, \cdots , n)$ with no underscoring, $(n - 1)n!/2$ decisions of the form $(\underline{1, 2}, 3, \cdots , n)$ with one pair of means underscored, $(n - 2)n!/3!$ decisions of the form $(\underline{1, 2, 3}, 4, \cdots , n)$ with three means underscored, $\cdots$, $(n - 2)n!$ decisions of the form $(\underline{1, 2}, 3, 4, \cdots , n)$ with two overlapping pair of means underscored, and so on through often large numbers of many forms finishing with one decision of the form $(\underline{1, 2, \cdots , n})$ in which all means are underscored with the one line. The underscoring has the same interpretation as before, for example $(\underline{1, 2, \cdots , n})$ is the decision that the means m_1 , m_2 , $\cdots$, m_n are not significantly different from one another.

The statements of the respective decisions may alternatively be made in terms of the true means, μ_1 , μ_2 , $\cdots$, μ_n . The statement, "m_i is significantly less than m_j ," is equivalent to the statement, "μ_i is less than μ_j ." Thus, the decision (1, 2, 3), for example, implies the acceptance of the hypothesis that $\mu_1 < \mu_2 < \mu_3$. The statement, "m_i and m_j are not significantly different," is equivalent to the statement "μ_i *is unranked relative to* μ_j ," where this is taken to mean that there is insufficient evidence to tell whether μ_i is less than, equal to, or greater than μ_j . Thus the decision (2, 1, 3), for example, consists of accepting the hypothesis that "$\mu_2 < \mu_1$, $\mu_2 < \mu_3$, but μ_1 is unranked relative to μ_3 ."

4. CONCEPTS OF POWER AND SIGNIFICANCE

4.1 *Power Functions.*

In analysing the power of these tests we are first faced with the difficulty that none of them, not even in the simplest case involving only two means, is a two-decision procedure, whereas a power function as defined by Neyman and Pearson (13) is strictly a two-decision-test concept.

In the three-decision test in the simplest case of two means, one way of avoiding this difficulty is to group the decisions (1, 2) and (2, 1) together as the decision that m_1 and m_2 are significantly different, or in other words as acceptance of the hypothesis $\mu_1 \neq \mu_2$. A convenient notation for this decision is $(1 \neq 2)$. The given three-decision test is reduced in this way to a two-decision procedure with decisions (1, 2) and $(1 \neq 2)$ and as such may be analysed as an α-level test of $\mu_1 = \mu_2$ against the two-sided alternative $\mu_1 \neq \mu_2$. The power function obtained in this way is given by the probability of the decision $(1 \neq 2)$ expressed as a function of the true difference $\epsilon = \mu_1 - \mu_2$. This may be conveniently denoted by $p(1 \neq 2)$, thus

$$p(1 \neq 2) = P[\text{dec. } (1 \neq 2) \mid \epsilon, \sigma^2].$$

An example of $p(1 \neq 2)$ is illustrated by the familiar curve shown by the dotted line in Figure 1b.

Although $p(1 \neq 2)$ is a most desirable function for measuring the properties of a test of $\mu_1 = \mu_2$ against $\mu_1 \neq \mu_2$ it has a serious weakness for measuring the properties of a three-decision test of two means. By pooling the probabilities of the two decisions (1, 2) and (2, 1) for any given value of the true difference, it combines the probability of the correct decision (that μ_1 or μ_2 is the higher mean as the truth may be), with the probability of the most incorrect decision (that μ_1 is the higher mean when in fact μ_2 is, or that μ_2 is the higher mean

when in fact μ_1 is). A function which combines probabilities of correct decisions with probabilities of serious errors in this way, is of no value in measuring desirable or undesirable properties. For this reason $p(1 \neq 2)$ will not be used as a measure of power in this problem. It has been discussed only because this function is so familiar that otherwise readers might have expected to have seen it used.

A more useful analysis of a three-decision test of two means is one which treats it as the joint application of two two-decision tests, namely, a test of the hypothesis, $\mu_1 \leq \mu_2$ against the alternative $\mu_2 < \mu_1$, and a test of the hypothesis $\mu_2 \leq \mu_1$ against the alternative $\mu_1 < \mu_2$. This type of analysis, which is suggested in a more general form by Lehmann (11, section 11), avoids the difficulties inherent in the $p(1 \neq 2)$ function, and extends readily to cases with more than two means.

From this point of view, a three-decision test has two power functions

$$p(2, 1) = P[\text{dec. } (2, 1) \mid \epsilon, \sigma^2]$$

and

$$p(1, 2) = P[\text{dec. } (1, 2) \mid \epsilon, \sigma^2],$$

which are the Neyman-Pearson power functions of the tests of $\mu_1 \leq \mu_2$ and $\mu_2 \leq \mu_1$ respectively. Examples of these functions are illustrated by the sigmoid and the reverse-sigmoid curves respectively in Figure 1b. Each of these functions has the merit that for any given value of the true difference ϵ, the function gives the probability of a correct *or* incorrect decision, and it is therefore clear whether the function should be as high or as low as possible. For example, $p(2, 1)$ represents the probability of deciding that μ_1 is the higher mean. Clearly then, it will be desirable for $p(2, 1)$ to be as high as possible for $\epsilon = \mu_1 - \mu_2 > 0$, and to be as low as possible for $\epsilon \leq 0$.

In the general case of n means we shall use $_nP_2$ power functions of the form

$$p(i, j) = P[\text{dec. } (i, j) \mid \mu_1, \mu_2, \cdots, \mu_n, \sigma^2],$$

where decision (i, j) includes all decisions which rank μ_i lower than μ_j; and $i, j = 1, 2, \cdots, n; i \neq j$. Each function $p(i, j)$ is the Neyman-Pearson power function of the test of the hypothesis $\mu_j \leq \mu_i$ against the alternative $\mu_i < \mu_j$. In general, therefore, $p(i, j)$ measures the probability of a correct decision with respect to μ_i and μ_j, over all values of the true means for which $\mu_i < \mu_j$, and the probability of a wrong decision over all values of the means for which $\mu_j \leq \mu_i$.

This approach is greatly simplified in all tests we wish to consider as a result of the reasonable symmetry restriction that all test properties be invariant under all $n!$ permutations of the true means. In other

words any test we consider must have the same properties for any set of values of the means irrespective of the identification of (the varieties represented by) the given means. Under these conditions it is necessary to investigate only one of the power functions $p(i, j)$ in order to investigate them all. An example of this is shown by the symmetry of $p(2, 1)$ and $p(1, 2)$ in Figure 1b.

4.2 *Significance Levels.*

So far as joint test properties are concerned only a relatively small number of significance levels need be considered. These are chosen so as to be as few in number as possible and yet have the property that once they are fixed at appropriate values, the merits of a test can then be judged solely in terms of its individual power functions.

In the simplest case involving only two means the significance levels or maximum type 1 error probabilities of the tests of $\mu_1 \leq \mu_2$ and $\mu_2 \leq \mu_1$ considered individually both occur when $\mu_1 = \mu_2$ and, by symmetry, these levels are equal. Because of this, only one significance level need be considered for the joint test, and this level may be taken as

$$\alpha = P[\text{dec. } (1 \neq 2) \mid \mu_1 = \mu_2],$$

which is the familiar significance level of the Neyman-Pearson test of $\mu_1 = \mu_2$ against $\mu_1 \neq \mu_2$. Given that α is fixed at α_0 the significance levels of the individual tests must be $\frac{1}{2}\alpha_0$ each.

In further discussion a type 1 error in a test of $\mu_i \leq \mu_j$, namely the decision (j, i) in cases where $\mu_i \leq \mu_j$, may be usefully termed an *error of wrong ranking* or the finding of a *wrong significant difference*. The importance of fixing α at α_0 may then be said to rest, not so much on the fact that the probability of a wrong ranking when $\mu_1 - \mu_2 = 0$ has been fixed at α_0, but on the fact that the probability of a wrong ranking at any value of the difference $\mu_1 - \mu_2$ cannot exceed α_0.

Any test for the case of three means may be regarded as having four significance levels of a nature similar to the significance level of a two-mean test. Three of these are of the form

$$\alpha(1, 2) = \text{maximum } P[\text{dec. } (1 \neq 2) \mid \mu_1 = \mu_2],$$

where the decision $(1 \neq 2)$ includes all decisions which rank μ_1 above or below μ_2 and the maximization is taken over all possible values of the true means μ_1, μ_2 and μ_3 for which $\mu_1 = \mu_2$. The level $\alpha(1, 2)$ is, moreover, the maximum value of the probability of making a wrong ranking of μ_1 and μ_2 over all possible values of the true means. The remaining two levels of this same form are

$$\alpha(1, 3) = \text{maximum } P[\text{dec. } (1 \neq 3) \mid \mu_1 = \mu_3],$$

$$\alpha(2, 3) = \text{maximum } P[\text{dec. } (2 \neq 3) \mid \mu_2 = \mu_3],$$

and are the maximum probabilities of making a wrong ranking between μ_1 and μ_3 and between μ_2 and μ_3 in a similar way.

The fourth significance level involves all three means and is defined as

$$\alpha(1, 2, 3) = P[\text{dec. } \overline{(1, 2, 3)} \mid \mu_1 = \mu_2 = \mu_3],$$

where the decision $\overline{(1, 2, 3)}$ includes all decisions which rank at least one pair of the means relative to one another. In other words, decision $\overline{(1, 2, 3)}$ includes all the 19 decisions previously listed except decision $(1, 2, 3)$. This three-mean significance level is simply the probability of finding at least one wrong significant difference between m_1, m_2 and m_3, that is, of making at least one wrong ranking of any pair of the true means μ_1, μ_2, and μ_3.

In the case of four means there are eleven significance levels which may be defined in a similar way. Six of these are two-mean significance levels of the form

$$\alpha(1, 2) = \text{maximum } P[\text{dec. } (1 \neq 2) \mid \mu_1 = \mu_2],$$

where, as before, the decision $(1 \neq 2)$ includes all decisions ranking μ_1 and μ_2 relative to one another, and the maximization is taken over all values of the means μ_1, μ_2, μ_3 and μ_4 for which $\mu_1 = \mu_2$. The remaining five two-mean significance levels defined in a similar way are $\alpha(1, 3)$, $\alpha(1, 4)$, $\alpha(2, 3)$, $\alpha(2, 4)$ and $\alpha(3, 4)$.

Four of the levels in this case are three-mean significance levels of the form

$$\alpha(1, 2, 3) = \text{maximum } P[\text{dec. } \overline{(1, 2, 3)} \mid \mu_1 = \mu_2 = \mu_3],$$

where the decision $\overline{(1, 2, 3)}$ includes all decisions which rank at least one pair of the means μ_1, μ_2 and μ_3 relative to one another, and where the maximization is taken over all values of the true means for which $\mu_1 = \mu_2 = \mu_3$. The remaining three three-mean significance levels similarly defined are $\alpha(1, 2, 4)$, $\alpha(1, 3, 4)$ and $\alpha(2, 3, 4)$.

Finally there is a single four-mean significance level defined as

$$\alpha(1, 2, 3, 4) = P[\text{dec. } \overline{(1, 2, 3, 4)} \mid \mu_1 = \mu_2 = \mu_3 = \mu_4],$$

where decision $\overline{(1, 2, 3, 4)}$ represents all decisions which rank at least one pair of the four means relative to one another. In other words decision $\overline{(1, 2, 3, 4)}$ includes all decisions except decision $(1, 2, 3, 4)$, which, following the previous pattern, is the decision that none of the differences among the four means is significant.

In a general test of n means, there are $_nC_2$ two-mean significance levels, $_nC_3$ three-mean significance levels, and so on up to $_nC_n = 1$ n-mean significance level. A p-mean significance level in general represents the maximum probability of finding at least one wrong significant difference among p observed means.

On careful consideration it appears that all* errors of wrong ranking in a test of n means can be adequately controlled by fixing these significance levels at appropriate values. The problem of finding a good test is then reduced to finding a procedure which optimizes the power functions $p(i, j)$ given that these significance levels are fixed at the chosen values.

4.3 *Protection Levels.*

The complement of any p-mean significance level may be termed a *p-mean protection level*, and is the minimum probability of finding no wrong significant differences among p observed means. The name "protection level" is suitable in that the level measures protection against finding wrong significant differences.

Thus, in a two-mean test, there is one protection level

$$\gamma = P[\text{dec. } \underline{(1, 2)} \mid \mu_1 = \mu_2] = 1 - \alpha.$$

If the significance level is 5%, for example, the protection level is 95%.

In a three-mean test, there are three two-mean protection levels $\gamma(1, 2)$, $\gamma(1, 3)$ and $\gamma(2, 3)$, where, for example,

$$\gamma(1, 2) = \text{minimum } P[\text{dec. } \underline{(1, 2)} \mid \mu_1 = \mu_2] = 1 - \alpha(1, 2)$$

and decision $\underline{(1, 2)}$ includes all decisions for which μ_1 and μ_2 are not ranked relative to one another. In addition there is one three-mean protection level

$$\gamma(1, 2, 3) = P[\text{dec. } \underline{(1, 2, 3)} \mid \mu_1 = \mu_2 = \mu_3] = 1 - \alpha(1, 2, 3).$$

In a general test of n means there are $_nC_p$ p-mean protection levels of the form

$$\gamma(a_1, a_2, \cdots, a_p)$$

$$= \text{minimum } P[\text{dec. } \underline{(a_1, a_2, \cdots, a_p)} \mid \mu_{a_1} = \mu_{a_2} = \cdots \mu_{a_p}]$$

where $p = 2, 3, \cdots, n$, each one being the complement of the corresponding significance level. The symbols $a_1, a_2, \cdots, a_p$ stand for the subscripts identifying the particular set of p means concerned.

---

*See also comments on class 2 protection levels in section 5.4.4.

(Thus decision $(a_1, a_2, \cdots, a_p)$ represents the decision that there are no significant differences between the observed means m_{a_1}, $m_{a_2}, \cdots, m_{a_p}$).

In further discussion of the controlling of errors of wrong ranking it will be somewhat more convenient to think in terms of fixing the protection levels of a test rather than in terms of fixing the significance levels.

4.4 *Consistent Protection Levels.*

We now consider the important question: In any test of n means, given that γ_2 is an appropriate value for the two-mean protection levels, what values $\gamma_3, \gamma_4, \cdots, \gamma_n$ should be regarded as satisfactory for the three-mean, four-mean, etc., protection levels, and for the n-mean protection level?

First it should be noted that if a symmetric test with optimum power functions were constructed subject only to a restriction on the value γ_2, the higher order protection levels would almost invariably be too low to be satisfactory. For example in the case of four means when $n_2 = \infty$, a test of this type with $\gamma_2 = 95\%$ would be obtained by applying six 5% level symmetric normal-deviate tests to each of the six differences between the four means. The four-mean protection level of this *multiple normal-deviate test*, as it may be termed, will be seen later to be only $\gamma_4 = 79.7\%$. That is, the minimum probability of finding no wrong significant differences between the four means. is only 79.7%. This is too low to be satisfactory. The three-mean protection levels in the same test have the value $\gamma_3 = 87.8\%$ which is also too low.

On the other hand, it does not necessarily follow that all of the higher order protection levels should be raised to the value γ_2 of the two-mean protection level as some writers have implicitly assumed. Any increases in the latter levels must necessarily be made at the expense of losses in power (that is, of increases in probabilities of type 2 errors), and it is most important that the levels be raised no more than is absolutely necessary. We shall now show that there are good reasons* for raising the higher order protection levels only part of the way towards the value of the two-mean protection levels.

Suppose, for the sake of an example, that a randomized block experiment were designed for the purpose of testing (a) the difference between two varieties V_1 and V_2, (b) the difference between two fertilizers F_1 and F_2 and (c) the difference between two insect control

---

*See also (5, section 6) and (6, p. 177).

spray methods S_1 and S_2. If interactions could be assumed to be zero, as might well be reasonable, a good design would be obtained by randomizing the four treatment combinations $V_1F_1S_1$, $V_1F_2S_2$, $V_2F_1S_2$ and $V_2F_2S_1$ within each block, where $V_1F_1S_1$, for example, denotes the application of fertilizer F_1 and spray method S_1 in a plot sown with variety V_1. If the observed means of these combinations are denoted respectively by m_1, m_2, m_3 and m_4, the varietal, fertilizer and spray differences would be measured respectively by the independent differences:

$$d_1 = (m_1 + m_2) - (m_3 + m_4) = m_1 + m_2 - m_3 - m_4$$

$$d_2 = (m_1 + m_3) - (m_2 + m_4) = m_1 - m_2 + m_3 - m_4$$

$$d_3 = (m_1 + m_4) - (m_2 + m_3) = m_1 - m_2 - m_3 + m_4$$

Now, provided that the number, r, of replications and hence the number of error degrees of freedom, $n_2 = 3r$, were large enough, it would be possible to make independent tests of the three given differences. Under these circumstances, if, say, a 5% level test of each difference were desired, no reasonable objection could be raised to the joint unmodified application of three 5% level tests. The joint use of these tests would be just as valid as if the differences were tested in three independent and separate experiments. In this joint test, it is clear that if the three null hypotheses in the individual tests were simultaneously true, which would imply that the true means μ_1, μ_2, μ_3, and μ_4 of the four combinations were all equal, the probability of not rejecting this joint hypothesis would be $(.95)^3 = 85.7\%$. Although this value is lower than 95%, it is clearly an implicitly unobjectionable result of having chosen a 95% protection level for each of the independent tests.

Now, the error of wrongly rejecting the hypothesis $\mu_1 = \mu_2 = \mu_3 = \mu_4$ in this type of test is no less serious than the error of rejecting the same hypothesis in the type of test under consideration, and a four-mean protection level is the probability of not making an error of this kind. Hence, it is argued that the objections to the low four-mean protection level $\gamma_4 = 79.7\%$ of the 5% level multiple normal-deviate test above would be appropriately remedied if the level were raised to $\gamma_4 = 85.7\%$.

A similar analogy with two independent 5% level tests of two independent differences among three means can be invoked for choosing an appropriate value for the three-mean protection levels in the same test. This leads to the conclusion that the objection to the low value $\gamma_3 = 87.8\%$ for these levels would be removed if they were increased to $(.95)^2 = 90.25\%$.

The same argument readily generalizes to give the result that the value $\gamma_p = \gamma_2^{p-1}$ for any p-mean protection level is appropriate in association with the value γ_2 for a two-mean protection level. The exponent $p - 1$ in these levels is given by the number of independent comparisons which can be specified, or the degrees of freedom, among the p means. For this reason the levels $\gamma_p = \gamma_2^{p-1}$ may be termed *protection levels based on degrees of freedom*.

Protection levels of this type have been used in constructing the multiple comparisons test (6, 7) and the new multiple range test. In the example of section 2 giving a 5% level new multiple range test of the seven barley variety means, the values of the protection levels are: $\gamma_2 = 95\%$, $\gamma_3 = 90.25\%$, $\gamma_4 = 85.7\%$, $\gamma_5 = 81.5\%$, $\gamma_6 = 77.4\%$ and $\gamma_7 = 73.5\%$. Since $\gamma_2 = 95\%$, we know that the probability of finding a significant difference between any two means when the corresponding true means are equal is definitely less than or equal to 5%. The higher order protection level values are in accord with this property.

In a similar 5% level test of 101 means, the first seven protection level values would be the same and the remainder would get progressively smaller down to $\gamma_{101} = (.95)^{100} = 0.6\%$ for the 101-mean protection level. Despite the independent tests analogy already given, the higher order protection levels may appear unduly low unless their progressively diminishing importance is fully realized. The appropriateness of these higher order protection levels in general will be emphasized by a further discussion of the independent tests analogy with particular reference to the justification of the 101-mean level $\gamma_{101} = 0.6\%$.

To take a corresponding analogy, suppose that in the course of a year's work, an experimenter has tested 100 separate null hypotheses H_1, H_2, $\cdots$, H_{100} in 100 independent experiments, and that he has chosen a 5% level test in each case. Should he be alarmed over the obvious fact that *if* the 100 null hypotheses were simultaneously true there has been only a 0.6% chance of not rejecting this joint hypothesis? Clearly the answer is no, because it would be illogical to alter any given individual test for reasons entirely independent of that test.

In choosing a 5% level of significance in each test the experimenter has implicitly expressed the opinion that there is some *a priori* chance that the respective null hypothesis is not true. It can be stated as a general rule that the more one can argue against the truth of a null hypothesis on *a priori* grounds the lower, other things being equal, should be the protection level of the test, in order not to waste power in detecting the truth of the alternative hypothesis. In choosing a 5% level test which has a 95% protection level the experimenter is implicitly prepared to assume that the *a priori* probability of the null

hypothesis is less than unity and lower than if, for example, he had chosen a 1% level test which has a 99% protection level.

Now, if the individual null hypotheses are independent in the sense that their *a priori* probabilities are independent, and if these probabilities are each appreciably less than unity as is implied by the choice of 5% levels of significance, the joint *a priori* probability for p such null hypotheses will be the product of the individual probabilities and will get less and less as p increases. Hence in the interests of not wasting power in detecting the truth of alternatives, it can well be appropriate to have lower and lower protection levels for each joint null hypothesis as p increases. In the case of the joint null hypothesis that all of the 100 individual null hypotheses are simultaneously true, for example, the *a priori* probability would be so small that it may be wasteful to use more than a very low protection level.

On extending this line of argument to a full average-weighted-risk analysis (24) including considerations of error weight functions and more complete Bayes (*a priori* probability) functions, the appropriateness of the overall joint test can be fully substantiated. In the full analysis the result is found to depend not directly on the independence of the Bayes functions of the individual tests, but on a closely related property, namely, the additivity of the error weight functions of the individual tests. An interesting more general form of this result, the proof and discussion of which will be presented subsequently as a separate paper, may be summarized as follows:

> Let T represent the joint test formed by k individual tests T_1, T_2, $\cdots$, T_k. Suppose that the error weight functions of the individual tests are additive in the sense that the error weight or loss for any joint decision D given any joint hypothesis H in the joint test T is equal to the sum of the error weights or losses for the decisions D_1, D_2, $\cdots$, D_k given the respective hypotheses H_1, H_2, $\cdots$, H_k, where the latter are individual test decisions and hypotheses forming D and H respectively.

> Then it follows, that if each individual test T_i is an optimum procedure from the point of view of minimizing average weighted risk, the joint test T is also an optimum procedure in the same sense.

Applying this to our example with 100 independent 5% level tests, we can say that since the error losses from one test to the next are additive, which is reasonable to assume because of the independent nature of the tests, and if each 5% level has been chosen as the best level to use for each test considered individually, then all features of

the joint test are optimum including, among many others, the low 0.6% protection level under special consideration.

A corresponding argument may be developed concerning the higher order protection levels in a test of the differences between n means. The larger the number of means involved, the less the *a priori* chance that the means will be homogeneous and the less, therefore, the need for a high protection level. The 101-mean protection level value of 0.6% in a 5% level multiple range test of 101 means, for example, may well be an optimum value for this level because of the remoteness of the possibility that all of the 101 true means are equal.

Owing to added complexities, it has not been possible thus far to prove in complete detail that protection levels based on degrees of freedom are exactly optimum in these tests also. However, since such protection levels are optimum in sets of independent tests, and since their functions are so similar in these tests, it is safe to conclude at least that they are close to optimum, and far closer than their only proposed rivals, namely, levels which are all equal to the two-mean protection level. It therefore seems sound practice to use these levels until they can be further improved by a more thorough minimum average risk analysis.

Having defined a set of relations among the values of the p-mean protection levels of a test, we therefore need to specify only one of these values and the remainder are fixed accordingly. From a practical point of view it is most pertinent and useful to define the levels in the way adopted in the multiple comparisons test (6, 7) and retained in the new multiple range test. The example given for the latter test in section 2 is a 5% level test in the sense that its two-mean significance levels are 5% and the protection levels are $\gamma_p = (.95)^{p-1}, p = 2, 3, \cdots, 7$. Likewise in a general test of n means, an α-level test denotes a procedure in which the two-mean significance levels are α and the protection levels are $\gamma_p = (1 - \alpha)^{p-1}, p = 2, 3, \cdots, n$. With the significance level of a test defined in this way, all that is necessary in choosing a level for a test of a given set of n means is to choose the level which would be considered appropriate for a test of the difference between any two of the means *assuming that the remaining means were not present*. Provided an appropriate value is chosen for this level, the remaining levels in the test are automatically fixed at their correspondingly appropriate values.

5. REVIEW OF SEVERAL TESTS

Comparisons will now be made between several test procedures which have been proposed for the given problem. In most of the detailed

discussion, consideration will be restricted to the following special simplifying conditions: The degrees of freedom for error will be assumed to be infinite, i.e., $n_2 = \infty$; the standard error of a mean will be assumed to be unity, i.e., $\sigma_m = 1$; and the significance level α of each test will be 5%, i.e., $\alpha = .05$. These will be referred to briefly as the special conditions $n_2 = \infty$, $\sigma_m = 1$ and $\alpha = .05$. This will provide a simple and familiar context for bringing out the main points of difference between the tests as clearly as possible. These main points are essentially unaltered when the special conditions are removed.

5.1 *The Symmetric Three-Decision t Test of Two Means.*

In the case of two means, the best test for choosing between the three possible decisions is the following familiar rule, which may be termed an *α-level symmetric three-decision t test*: *Make the decision* (1, 2) *if $m_1 - m_2 < -\sqrt{2}t_\alpha s_m$, the decision* (1, 2) *if $| m_1 - m_2 | \leq \sqrt{2}t_\alpha s_m$, or the decision* (2, 1) *if $m_1 - m_2 > \sqrt{2}t_\alpha s_m$*; where t_α is the two-tail α-level significant value of t.

Under the special conditions $n_2 = \infty$, $\sigma_m = 1$, $\alpha = .05$, the test reduces to a 5% *level symmetric three-decision normal-deviate test* and the significant difference $\sqrt{2}t_\alpha s_m = \sqrt{2}u_\alpha \sigma_m$ is the familiar value $1.960\sqrt{2} = 2.77$.

This test is satisfactory for the case of two means, and it is only when we pass on to consider tests involving more than two means that the differences arise in proposed test procedures. It is worthwhile, however, to consider various special details of an analysis of the three-decision normal-deviate test as an introduction to methods of analysing the more complex tests.

(i) *Sample Space.* A common useful method for representing this test graphically is shown in Figure 1a. In this figure, the horizontal straight line provides an example of a one-dimensional *sample space* and is used for plotting the observed difference $x = m_1 - m_2$. Any point on this line representing an observed value of x is called a *sample point*. The line is divided into three intervals, $x < -2.77$, $-2.77 \leq x \leq 2.77$, and $2.77 < x$. These represent the respective sets of points for which the decisions (1, 2), (1, 2) and (2, 1) are made and are termed *decision regions*. It is convenient to denote each region by the same symbol, (1, 2), (1, 2) or (2, 1), that is used for the corresponding decision.

(ii) *Parameter Space.* The straight line in Figure 1a may also be used for plotting values of the "true" difference, $\epsilon = \mu_1 - \mu_2$, between the true means involved. When used in this way, the line provides an example of a *parameter space*, as distinct from its function as a *sample*

space when used for plotting x. Any point on the line representing a given value of ϵ is called a *parameter point*.

(iii) *Probability Density.* In the special case we are considering, the probability distribution function $f(x; \epsilon)$ of a sample point x (ob-
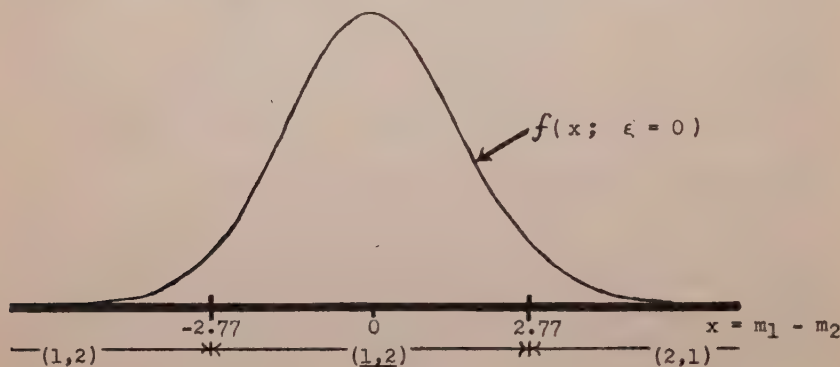
FIGURE 1a

Regions for a 5%-level symmetric three-decision normal-deviate test ($\sigma_x = \sqrt{2}$)

FIGURE 1b

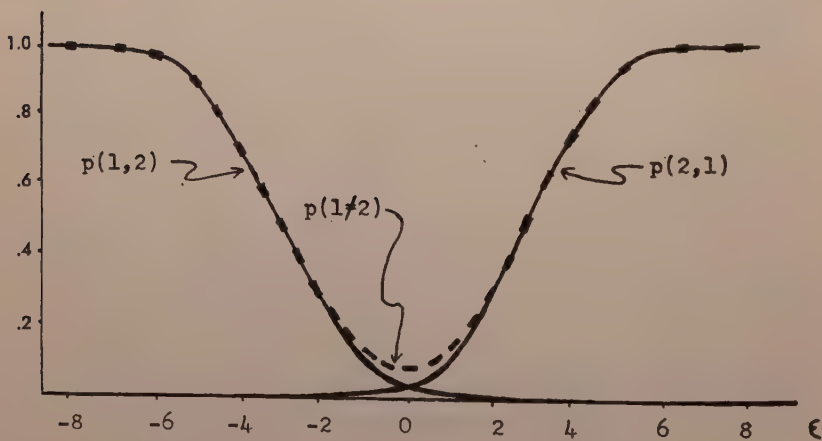Power Functions for 5% Level Symmetric Three-Decision Normal-Deviate Test ($\sigma_x = \sqrt{2}$)

served difference) about a given parameter point ϵ (given true difference) is given by a normal probability density function with mean ϵ and variance 2. For example, when $\epsilon = 0$ this function may be represented by the familiar curve shown in Figure 1a. The curve for any other value

of ϵ has the same shape and is located with its center over the given ϵ value.

(iv) *Power Functions.* The power function $p(1, 2)$ representing the probability of decision $(1, 2)$ for any given value of ϵ is given by the area under the probability density curve for the given ϵ, over the region $(1, 2)$. Likewise the power function $p(2, 1)$ for the same ϵ value is given by the area under the same curve and over the region $(2, 1)$. The functions $p(1, 2)$ and $p(2, 1)$ are represented by the reverse-sigmoid and the sigmoid curves in Figure 1b.
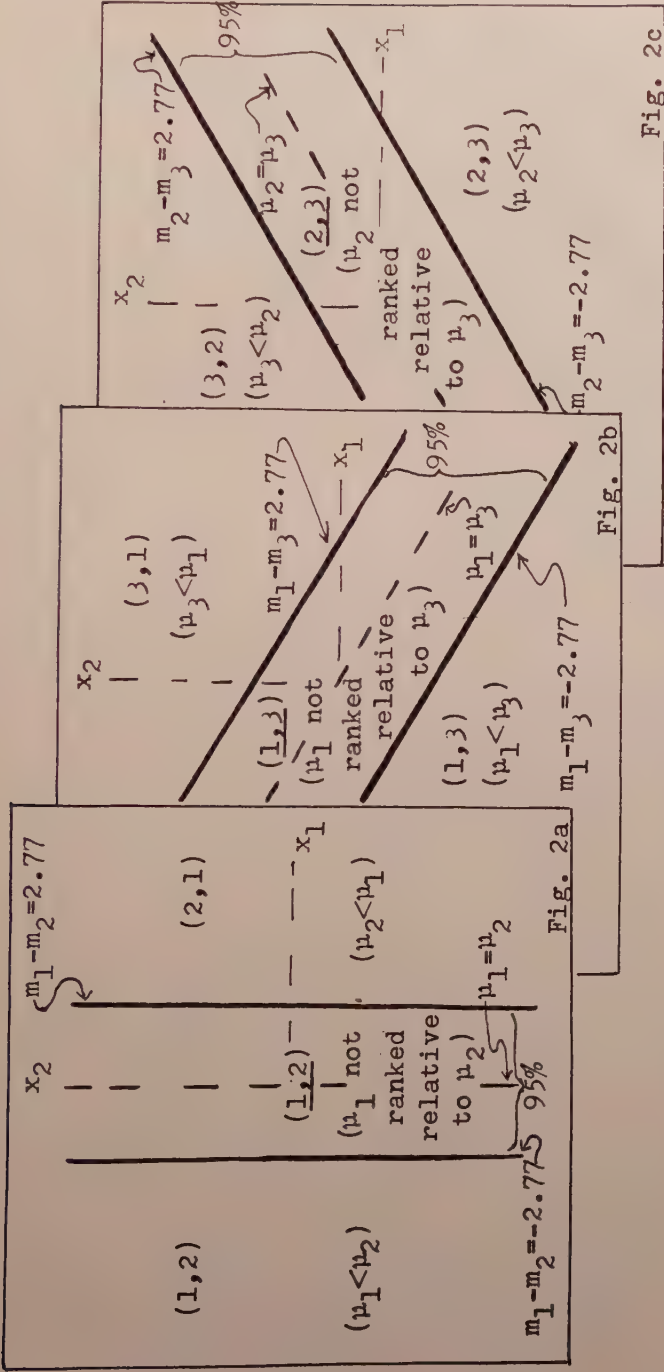
(v) *Significance and Protection Levels.* The significance level, $\alpha = 5\%$, of this test is represented by the sum of the ordinates of the power curves in Figure 1b at $\epsilon = 0$, each of which is $2\frac{1}{2}\%$. The protection level is $1 - \alpha = 95\%$. In Figure 1a, the significance level is the sum of the areas under the dotted curve for $\epsilon = 0$, over the regions $(1, 2)$ and $(2, 1)$. The protection level is the area of the same curve over the region $(1, 2)$. Extensions of these familiar ideas will be useful in illustrations of corresponding features in tests of more than two means.

The virtues of the 5% level normal-deviate three-decision test can be summarized most usefully as follows: The minimum protection against making a wrong ranking of the two means is 95%, and, for all procedures for which this is true, the power curves of this test are uniformly maximized over all values of ϵ for which they measure probabilities of correct decisions, and are uniformly minimized over all values of ϵ for which they measure probabilities of incorrect decisions. This provides a good example of the general usefulness of the new multiple power function analysis which we have adopted for this and for the more complex procedures.

5.2 *Tests of Three Means. General Details.*

(i) *Sample Space.* To represent a test involving three means, m_1, m_2, and m_3, a two-dimensional sample space or plane is required in place of the one-dimensional sample space or line used above for a two-mean test. In this two-dimensional space it is convenient to plot the difference $x_1 = m_1 - m_2$ on the horizontal axis and the comparison $x_2 = (m_1 + m_2 - 2m_3)/\sqrt{3}$ on the vertical axis as rectangular Cartesian coordinates. Figures 2, 2a, 2b and 2c, and all subsequent sample space illustrations use these particular coordinates. It will be noted that x_2 is distributed independently of x_1 and has the same variance, $\sigma_x^2 = 2\sigma_m^2$. This leads to certain helpful features of symmetry which will become evident as we proceed.

Any set of values for the three differences $m_1 - m_2$, $m_1 - m_3$, and $m_2 - m_3$, between the three means, can be represented by a sample

point (x_1, x_2) in this two-dimensional sample space. For example, the set of differences $m_1 - m_2 = 4$, $m_1 - m_3 = -1$, and $m_2 - m_3 = -5$, found in the sample of means $m_2 = 10$, $m_1 = 14$, $m_3 = 15$, gives $x_1 = 4$ and $x_2 = -2\sqrt{3}$. These differences would thus be represented by the point $(4, -2\sqrt{3})$ located 4 units to the right of and $2\sqrt{3}$ units below the center of the space. The inverse relations by which the differ-

FIGURE 2

Regions of 5% Level Multiple Normal-Deviate Test ($n_2 = \infty$, $\sigma_m = 1$)

ences can be obtained from a sample point are $m_1 - m_2 = x_1$, $m_1 - m_3 = (x_1 + \sqrt{3}x_2)/2$, and $m_2 - m_3 = (-x_1 + \sqrt{3}x_2)/2$. Thus a point $(-2, 1)$ represents the set of differences $m_1 - m_2 = -2$, $m_1 - m_3 = -(2 - \sqrt{3})/2$, and $m_2 - m_3 = (2 + \sqrt{3})/2$.

(ii) *Parameter Space.* The plane used as a sample space in these figures may also be used for plotting values of the "true" comparisons $\epsilon_1 = \mu_1 - \mu_2$ and $\epsilon_2 = (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{3}$ between the true means involved. When used in this way it is termed a *parameter space*, and values for ϵ_1 and ϵ_2 constitute a *parameter point* (ϵ_1, ϵ_2). In the parameter space we shall need to make frequent references to the parameter point $(\epsilon_1, \epsilon_2) = (0, 0)$, the origin, at which all true means are equal, i.e., at which $\mu_1 = \mu_2 = \mu_3$. Similarly we shall need to refer to the

FIGURES 2a, 2b, 2c

Development of Regions in Figure 2

dotted lines labelled $\mu_1 = \mu_2$, $\mu_1 = \mu_3$, and $\mu_2 = \mu_3$ in Figures 2a, 2b, and 2c, representing all points for which $\mu_1 = \mu_2$, $\mu_1 = \mu_3$, and $\mu_2 = \mu_3$, respectively. The position of a parameter point on any one of the lines depends on the magnitude of the third mean relative to the two equal means represented by the line.

(iii) *Probability Density.* The probability distribution of a sample point (x_1, x_2) depends only on (ϵ_1, ϵ_2) and from the definition of x_1 and x_2 it is readily seen that the distribution function $f(x_1, x_2; \epsilon_1, \epsilon_2)$ is a bivariate normal one. Each x_i is distributed normally and independently about ϵ_i as mean and with a variance of 2. The distribution for any parameter point (ϵ_1, ϵ_2) can be visualized geometrically as a bell-shaped surface standing on the sample space plane with its center located over the given parameter point.

5.3 *The Multiple t Test.*

To illustrate the way in which a test can be represented in the sample space, we shall consider a previously mentioned special case of the procedure obtained by applying an α-level symmetric three-decision t test separately to each of the hypotheses, $\mu_1 = \mu_2$, $\mu_1 = \mu_3$, and $\mu_2 = \mu_3$. This may be termed an *α-level multiple t test*, and readily generalizes to the case of n means in which the individual t tests are applied to all $_nC_2$ hypotheses of the form $\mu_i = \mu_j$ which equate the means considered in all possible pairs.

As has been pointed out, this procedure does not provide a satisfactory test for our problem, and it is definitely not recommended for this purpose. We use it here and at other points in the discussion because of the excellent introduction it affords to better but more complex procedures.

Under the special conditions $n_2 = \infty$, $\sigma_m = 1$, $\alpha = .05$, the α-level multiple t test reduces to the 5% level multiple normal-deviate test. The 19 regions of this test are as shown in Figure 2.

(i) *Decision Regions.* The regions of the joint test are formed by the symmetrical intersection of three sets of two-mean test regions as shown in Figures 2a, 2b, and 2c. In Figure 2a the lines $m_1 - m_2 = -2.77$ and $m_1 - m_2 = 2.77$ divide the sample space into three regions (1, 2), (1, 2), and (2, 1). The region (1, 2) consists of the entire vertical strip passing down the center of the plane between the lines $m_1 - m_2 = -2.77$ and $m_1 - m_2 = 2.77$. The regions (1, 2) and (2, 1) are the remainders of the sample space plane lying to the left and right of (1, 2), respectively. These are the regions of the test of $\mu_1 = \mu_2$ and are two-dimensional extensions of the corresponding one-dimensional regions in Figure 1a. The notation has the same meaning as before;

for example, if a point falls in $(1, 2)$ the decision $(1, 2)$ is made, namely that m_1 is significantly less than m_2.

Likewise, the lines $m_1 - m_3 = \pm 2.77$ in Figure 2b divide the sample plane into the three regions $(1, 3)$, $(\underline{1, 3})$, $(3, 1)$ for the test of $\mu_1 = \mu_3$, and the lines $m_2 - m_3 = \pm 2.77$ in Figure 2c divide the sample plane into the three regions $(2, 3)$, $(\underline{2, 3})$, $(3, 2)$ for the test of $\mu_2 = \mu_3$. The sets of regions for each of these tests are identical with those for the test of $\mu_1 = \mu_2$, except for a rotation about the origin which is $60°$ counterclockwise for the first and $60°$ clockwise for the second.

Each of the 19 product regions for the joint test in Figure 2 corresponds to one of the 19 decisions previously listed for the case of three means. For example, in the intersection of $(1, 2)$, $(3, 1)$, and $(3, 2)$ in the top left-hand corner of the figure, the associated decisions $(1, 2)$, $(3, 1)$, and $(3, 2)$ constitute the joint decision $(3, 1, 2)$. This, it will be recalled, is the decision that m_1 is significantly less than m_2, m_3 is significantly less than m_1, and m_3 is significantly less than m_2. The region involved may be thus conveniently denoted as the region $(3, 1, 2)$. Likewise the intersection of the regions $(\underline{1, 2})$, $(\underline{1, 3})$, and $(\underline{2, 3})$ is the hexagonal region at the center in which the decision $(\underline{1, 2, 3})$ is made. This may accordingly be denoted as the region $(\underline{1, 2, 3})$.

(ii) *Power Functions.* The power function $p(1, 2)$, to take one of the six power functions involved, may be visualized as a *power surface* $P[\text{dec. } (1, 2) \mid \epsilon_1, \epsilon_2]$ above the parameter space. The ordinate of the surface at any point (ϵ_1, ϵ_2) is given by the integral over the region $(1, 2)$ of the bell-shaped distribution for that point. Since the boundary of region $(1, 2)$ is parallel to the ϵ_2 axis it is clear that sections of the power surface for different values of ϵ_2 are identical. Each section is depicted by the reverse-sigmoid $p(1, 2)$ curve shown for the two-mean test in Figure 1b.

The remaining power functions $p(1, 3)$, $p(2, 3)$, $p(2, 1)$, $p(3, 1)$ and $p(3, 2)$ may be visualized as power surfaces, identical with the surface for $p(1, 2)$, except that the one for $p(1, 3)$ is rotated $60°$ counterclockwise about the origin, the one for $p(2, 3)$ is rotated a further $60°$ counterclockwise about the origin, and so on.

(iii) *Protection Levels.* The two-mean protection level $\gamma(1, 2) =$ minimum $P[\text{dec. } (\underline{1, 2}) \mid \mu_1 = \mu_2]$ is the minimum integral over the strip-region $(\underline{1, 2})$, of any of the normal bivariate distributions centered on the line $\mu_1 = \mu_2$. Since the boundaries of $(\underline{1, 2})$ are parallel to the line $\mu_1 = \mu_2$, the minimum is given by the integral for any one parameter point $(0, \epsilon_2)$, and is 95%. The remaining two-mean protection levels $\gamma(1, 3)$ and $\gamma(2, 3)$ can be seen to be 95% in the same way.

The only remaining protection level is the three-mean level

$\gamma(1, 2, 3) = P[\text{dec. } (\underline{1, 2, 3}) \mid \mu_1 = \mu_2 = \mu_3]$. This is given by the integral over the hexagonal region $(\underline{1, 2, 3})$ of the bell-shaped bivariate normal distribution centered at the origin $(0, 0)$. Since this region is the locus of all points for samples in which the range is less than 2.77, it follows that the integral is the probability $P[q_3 < 2.77]$, where q_p is the standardized range of a sample of p independent observations from a normal population. Tables for these probabilities are given by Pearson and Hartley (15), and from these a value of 87.8% is found for this three-mean protection level. According to the principle of protection levels based on degrees of freedom, the three-mean protection level should be 90.25%.

In the test of four means the twelve power functions are similar to those of the simpler cases in that $p(1, 2)$, for example, can be expressed as a function of $\mu_1 - \mu_2$ alone. In the reduced form $p(1, 2)$ is identical with the $p(1, 2)$ function of the two-mean test illustrated in Figure 1b. The six two-mean and four three-mean protection levels in this test are readily seen to be $P[q_2 \leq 2.77] = 95\%$ and $P[q_3 \leq 2.77] = 87.8\%$ as for the corresponding levels in the three-mean test. The four-mean protection level is similarly found to be $P[q_4 \leq 2.77] = 79.7\%$.

As has been mentioned previously, it is the lowness of the three-mean and four-mean protection levels in these tests which invalidates them as satisfactory 5% level procedures. On the other hand their power functions considered individually have all of the optimum properties of those of the two-mean test. Similar properties are possessed by α-level multiple t tests in general.

The general problem of finding a satisfactory test may be regarded as that of raising the higher order protection level values of an α-level multiple t test to acceptable values, by methods which interfere as little as possible with its optimum power functions.

5.4 *Multiple Range Tests.*

5.4.1 *The Newman-Keuls Test.*

A test proposed by Newman* (12) in 1939 and again by Keuls (10) in 1952 succeeds very simply in raising all of the low protection levels of the multiple t test. This test is equivalent to a multiple t test preceded by several preliminary range tests. Since the t tests of which the multiple t test is composed may be regarded as range tests of

---

*Newman mentions that the principle of this test was initially suggested to him by "Student."

subsets of two means each, the overall procedure is composed entirely of range tests and may be usefully termed a *multiple range test*.

An α-level Newman-Keuls multiple range test is given by the rule: *The difference between any two means in a set of n means is significant provided the range of each and every subset which contains the given two means is significant according to an α-level range test.* Thus in the case of three means under the special conditions $n_2 = \infty$, $\sigma_m = 1$, $\alpha = .05$, the difference $m_1 - m_2$, for instance, is significant when the range of m_1, m_2, m_3 exceeds 3.32 (the 5% level value of the range of three means) and $m_1 - m_2$ exceeds 2.77. In the case of four means, $m_1 - m_2$ is significant when the range of m_1, m_2, m_3, m_4 exceeds 3.63 (the 5% level value of the range of four means), the ranges of m_1, m_2, m_3 and m_1, m_2, m_4 each exceed 3.32, and $m_1 - m_2$ exceeds 2.77.
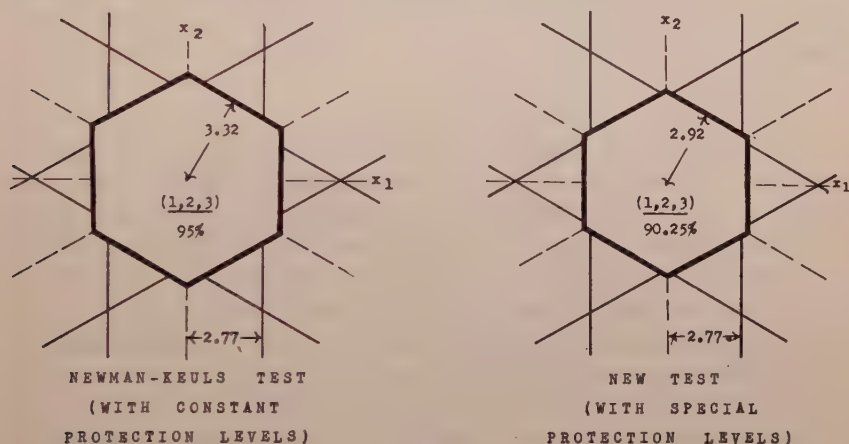

FIGURE 3

5% level multiple range tests ($n_2 = \infty$, $\sigma_m = 1$)

The regions of the three-mean test are shown in Figure 3. These are the same as those of the corresponding multiple normal-deviate test except for the changes caused by the expansion of the region (1, 2, 3) from a regular hexagon with radius* 2.77 to a regular hexagon with radius 3.32. This raises the three-mean protection level from 87.8% to 95%. On the other hand, the two-mean protection levels remain unaltered at 95%. For example, the level $\gamma(1, 2)$, which is the minimum integral over the modified strip region (1, 2) of any distribution

*The radius of a hexagon will be used as short for the radius of the *inscribed* circle of the hexagon.

centered on the line $\epsilon_1 = \mu_1 - \mu_2 = 0$, is unchanged because the region $(\underline{1, 2})$ is unaltered away from the origin $(\epsilon_1, \epsilon_2) = (0, 0)$. The integrals are larger than 95% at the origin but drop to 95% as $|\epsilon_2|$ increases.

The six power functions are readily seen to be similar to those of the corresponding multiple normal-deviate test except for a general lowering in the area around the origin. For example, $p(1, 2)$ which is the integral over the region $(1, 2)$ of the distribution centered at any point (ϵ_1, ϵ_2) is reduced by an amount equal to the integral over the trapezium shaped region which has been taken from $(1, 2)$ and added to $(\underline{1, 2})$. This reduction is greatest for a distribution centered at $(\epsilon_1, \epsilon_2) = (-3.04, 0)$ (the center of the trapezium) and gets less as the distance from this point increases.

In the test of four means, the four-mean and three-mean protection levels are raised from 87.8% and 79.7% respectively to 95%, and corresponding reductions in the power functions accompany these changes.

5.4.2 *The New Multiple Range Test.*

The new multiple range test applied to the barley yield data in section 2 is a multiple range test like the Newman-Keuls procedure, except that, as has already been emphasized, it employs the special protection levels system based on degrees of freedom. A general α-level multiple range test of this type is given by the rule: *The difference between any two means in a set of n means is significant provided the range of each and every subset which contains the given means is significant according to an α_p-level range test where $\alpha_p = 1 - \gamma_p$, $\gamma_p = (1 - \alpha)^{p-1}$, and p is the number of means in the subset concerned.*

Figure 3 shows the regions of this test applied to three means under the same special conditions as before. These regions are identical with those of the corresponding Newman-Keuls test, also shown in Figure 3, except that the center hexagon has a radius of 2.92 instead of 3.32 and the adjacent regions are changed accordingly. This is sufficient to give the test a three-mean protection level of 90.25%. The two-mean protection levels remain unaltered at 95%, the same as in the Newman-Keuls test.

The power functions of this test are similar to those of the Newman-Keuls test except that the reductions relative to the multiple normal deviate test are uniformly smaller, making the test uniformly more powerful. The reductions in $p(1, 2)$, for example, are given as before by integrals over the trapezium formed by the intersection of the center hexagon $(\underline{1, 2, 3})$ with the original $(1, 2)$ region in Figure 2a. Since the hexagon is smaller than in the previous test, the trapezium

is smaller, and the reduction integrals are therefore uniformly decreased. The difference in power is greatest at a point near the center $(-3.04, 0)$ of the bigger trapezium and diminishes towards zero with increase of distance away from this point.

In the case of four means, this test raises the four-mean protection level from 79.7% to 85.7% and the three-mean levels from 87.8% to 90.25% in a similar way. The two-mean protection levels remain unaltered at 95%. Likewise the power functions are uniformly lower than those of the corresponding multiple *t* test but uniformly higher than those of the corresponding Newman-Keuls test.

The gains in power in the new multiple range test are quite appreciable, expecially for some parameter points and are entirely due to use of protection levels based on degrees of freedom. In passing, the independent tests analogy used in support of these new levels may be illustrated for purposes of comparison by the regions of the test shown in Figure 4. These are the regions of two 5% level independent normal deviate tests of $x_1 = m_1 - m_2$ and $x_2 = (m_1 + m_2 - 2m_3)/\sqrt{3}$ respectively, assuming $n_2 = \infty$ and $\sigma_m = 1$ as before. Tests like these would be needed, for example, if m_1 and m_2 were grain yields from two strains of one barley variety (A) and m_3 were the yield of another variety (B). Attention under these circumstances might well be restricted to testing the difference x_1 between the two strains of variety A and the difference x_2 between the two varieties A versus B.

The case for protection levels based on degrees of freedom may be put very briefly in terms of the tests illustrated in Figures 3 and 4, as follows: Because of the independence of its two component tests, the joint test in Figure 4 is a valid and acceptable joint procedure. The square region $(1, 2, 3)$ at the center of this test has the same function as the hexagonal region at the center of a multiple range test in that it is the locus of all points which do not lead to the rejection of the hypothesis $\mu_1 = \mu_2 = \mu_3$ (which implies $(\epsilon_1, \epsilon_2) = (0, 0)$). It is adequate, therefore, to increase the dimensions of the hexagonal region in a multiple range test only so far as is needed to make the integral of the distribution at origin $(0, 0)$ over this region equal to the integral of the same distribution over the square region in Figure 4. The latter integral is 90.25% and the hexagonal region of the new multiple range test in Figure 3 has been constructed in this way.

5.4.3 *Tukey's Test Based on "Allowances."*

In 1951 Tukey (22) introduced a procedure for estimating confidence intervals, or "allowances" as he called them, for the differences $\mu_i - \mu_j$ which we have been considering. He defined a confidence coefficient

β for the joint procedure as the probability that all intervals simultaneously contain the values of the corresponding true differences. This method can be used to give, among other things, a significance test for our general problem. If, in a procedure with confidence coefficient β, the confidence interval for $\mu_i - \mu_j$ is denoted by $I_{ij}(\beta)$ this test may be expressed as the following rule: *Make the decision* (i, j) *if*
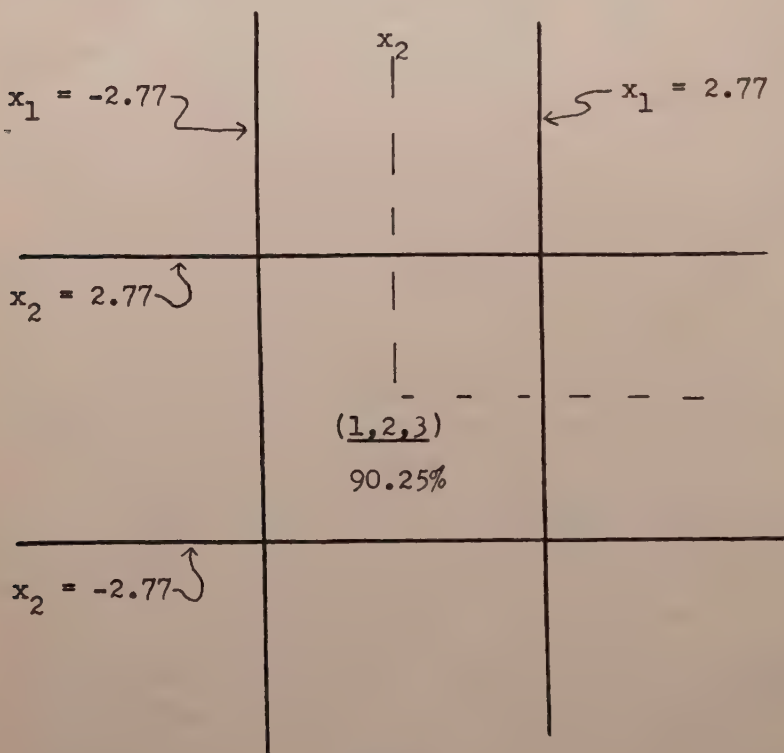

$x_1 = -2.77$

$x_1 = 2.77$

x_2

$x_2 = 2.77$

$x_2 = -2.77$

$(\underline{1,2,3})$

90.25%

FIGURE 4

Regions for 5% Level Joint Normal-Deviate Tests of Two Independent Comparisons ($n_2 = \infty$, $\sigma_x = \sqrt{2}$)

$I_{ij}(\beta)$ *lies to the left of zero, the decision* $(\underline{i, j})$ *if* $I_{ij}(\beta)$ *includes zero, or the decision* (j, i) *if* $I_{ij}(\beta)$ *lies to the right of zero.* An α-level test, by the originator's definition, is obtained by putting $\beta = 1 - \alpha$.

The test given in this way for three means, under the special conditions $n_2 = \infty$, $\sigma_m = 1$, $\alpha = .05$, is identical with the multiple normal-deviate test shown in Figure 2 except that the width of each of the strips $(\underline{1, 2})$, $(\underline{1, 3})$, $(\underline{2, 3})$ is increased from 2×2.77 to 2×3.32. The

method of derivation from confidence intervals implicitly imposes the restriction that the boundaries of $(1, 2)$, $(1, 3)$, and $(2, 3)$ must be parallel straight lines. The distance between the lines is widened until the dimensions of the center hexagon $(1, 2, 3)$ are as large as those of the Newman-Keuls test, thus making the three-mean protection level $1 - \alpha = 95\%$. At the same time the two-mean protection levels are increased uniformly from 95% to 98.1%. This test is readily seen to be more conservative and uniformly less powerful than any of the previous procedures.

5.4.4 *Tukey's 1953 Multiple Range Test.*

In 1953 Tukey (23) relaxed the conservatism of the previous test somewhat by proposing a multiple range procedure in which the significant ranges are each midway between the ones required by the test based on allowances and those required by the Newman-Keuls test. In the case of three means, under the same special conditions as before, the regions of this test are the same as those of the Newman-Keuls procedure except that the widths between the parallel lines are increased from 2.77 to $\frac{1}{2}(2.77 + 3.32) = 3.04$. The hexagon radius is 3.32 in both tests.

In suggesting this test, Tukey drew attention to an important point which may be illustrated by the following example. Suppose that in a 5% level Newman-Keuls test of four means, again assuming $n_2 = \infty$ and $\sigma_m = 1$, the values of the true means are $\mu_1 = \mu_2 = \mu$ and $\mu_3 = \mu_4 = \mu + \delta$. Suppose the difference δ between the two groups of means is so large that the preliminary range tests are practically certain to be significant, then the probability of jointly deciding that both $|m_1 - m_2|$ and $|m_3 - m_4|$ are not significant is $P[|m_1 - m_2| \leq 2.77] \times P[|m_3 - m_4| \leq 2.77] = 90.25\%$. This is an example of a whole set of levels, which we may call class 2 protection levels, which are not raised to $(1 - \alpha)$ in an α-level Newman-Keuls test and are more akin to levels based on degrees of freedom. Both of Tukey's procedures have been designed with the objective of raising these class 2 protection levels along with the others to at least $(1 - \alpha)$. The 1953 test is a modification of the test based on allowances which is uniformly more powerful than the later but which, Tukey judges, still meets his given objective.

When protection levels based on degrees of freedom are adopted, as in the new multiple range test, the class 2 levels are automatically fixed at, or slightly above (when n_2 is small), their appropriate values and need no special attention.

In the case of the Newman-Keuls procedure it is not clear whether

either one of the authors was aware of the presence of these lower levels and whether he would wish to defend them as this writer does or not.

5.5 *Multiple F Tests.*

A series of tests paralleling the above multiple range tests can be defined using F tests instead of range tests. These may conveniently be termed multiple F tests. Thus, corresponding to the new multiple range test, an α-*level multiple F test with protection levels based on degrees of freedom* may be defined by the following rule: Rule 1. *The difference between any two means in a set of n means is significant provided the variance of each and every subset which contains the given means is significant according to an α_p-level F test where $\alpha_p = 1 - \gamma_p$, $\gamma_p = (1 - \alpha)^{p-1}$, and p is the number of means in the subset concerned.*

In the case of three means under the special conditions $n_2 = \infty$, $\sigma_m = 1$, $\alpha = .05$, the regions of this test are as shown in Figure 5. These regions are the same as those of the corresponding multiple normal-
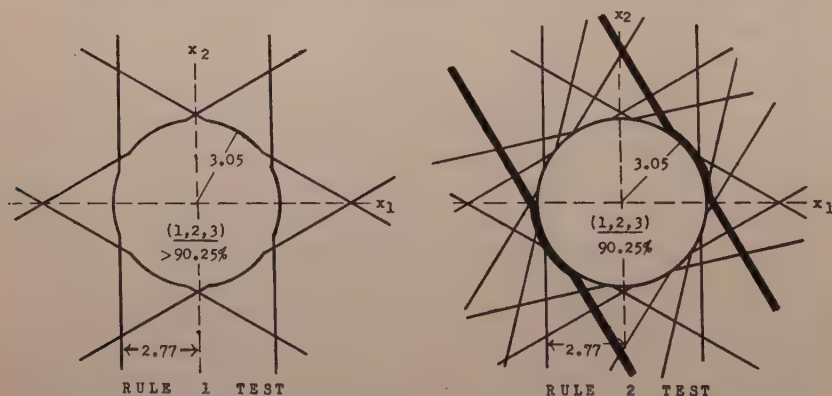

FIGURE 5

5% level multiple F tests with special protection levels ($n_2 = \infty$, $\sigma_m = 1$)

deviate test except that the strip-regions $(1, 2)$, $(1, 3)$, $(2, 3)$ have their boundaries expanded to those of the circle centered at the origin, with radius 3.05. This radius 3.05 is calculated as $\sqrt{4F}$, where* F is the 9.75% significant value of an F ratio with degrees of freedom 2 and ∞. If the center region $(1, 2, 3)$ were comprised of the circle alone, this would raise the three-mean protection level to just 90.25%

---

*This test requires special F tables or equivalent tables as given in (6), Tables 1 and 2.

as desired. The six small areas outside the circle but inside $(\underline{1}, \underline{2}, \underline{3})$ give the test a slightly higher protection level than 90.25%, which is not necessary and makes some modification of Rule 1 desirable.

The multiple F test can be generalized to test the significance of all linear comparisons of the form $c = \sum_{i=1}^{n} k_i m_i$, where $k_1, k_2, \cdots, k_n$ is any set of arbitrary constants such that $\sum_{i=1}^{n} k_i = 0$. (Each linear function of this form can be regarded as the difference between weighted means of two subsets of the full set of means.) The general rule is: Rule 2. *Any comparison of the form* $c = \sum_{i=1}^{n} k_i m_i$ *is significantly different from zero provided the variance of each and every subset which contains all of the means involved in c is significant according to an α_p-level F test and provided also that c differs significantly from zero according to an α-level t test where $\alpha_p = 1 - \gamma_p$, $\gamma_p = (1 - \alpha)^{p-1}$, and p is the number of means in the subset concerned.* By "all of the means involved in *c*" is meant all means which have non-zero coefficients in the linear function $c = \sum_{i=1}^{n} k_i m_i$.

The regions of this more general test, under the same special conditions, are also shown in Figure 5. The three intersecting strip regions given by Rule 1 are now replaced by an infinity of strips, all of which pass symmetrically through the center of the sample space and intersect each other at all angles. Each strip and the areas to either side of it represent the test regions for the comparison measured at right angles to the axis of the strip. For example, the strip region between the heavy lines in the illustration contains points for samples in which the comparison $c = \frac{1}{2}m_2 + \frac{1}{2}m_3 - m_1$ is not significantly different from zero. The areas to either side of this region contain points for samples in which the comparison is significantly positive or negative.

5.5.1 *The Multiple Comparisons Test.*

The *multiple comparisons test* proposed by the author in 1951 (6, 7) is a multiple F test which consists of a compromise between Rule 1 and Rule 2. As many significant differences as possible are found by the Rule 1 test. Rule 2 is then used to test any comparisons of interest within subsets of means not already found to contain significant differences by Rule 1.

Figure 6 shows the regions of this test under the same special conditions as before. These regions are identical with those of the Rule 1 test in Figure 5 except for the additional six regions lying outside the circle and inside the original hexagon. These represent regions in which comparisons involving all three means are found to be significant. In the small region at the top of the circle, for example, various weighted means of m_1 and m_2 are significantly larger than m_3.
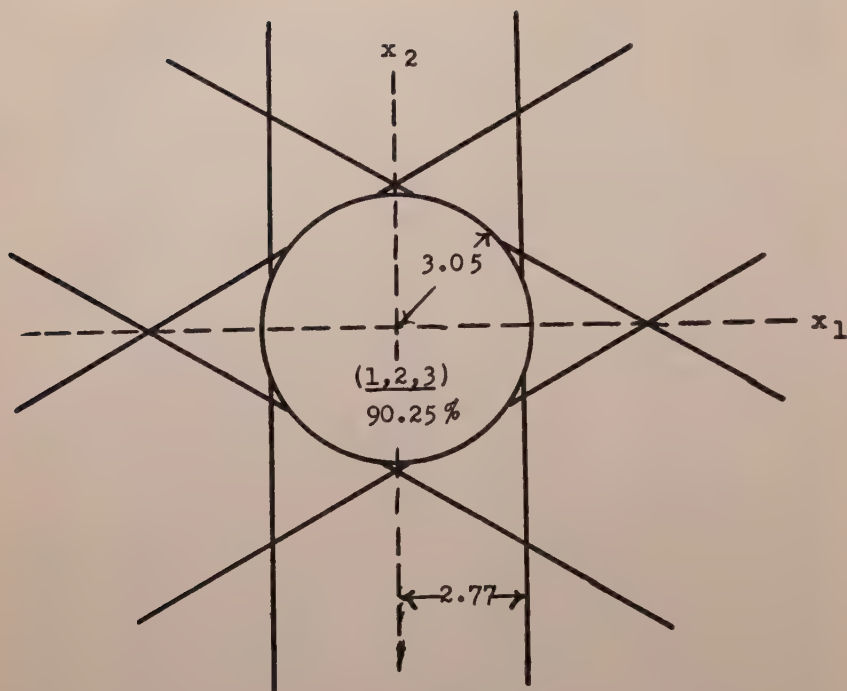
FIGURE 6

5% Level Multiple Comparisons Test ($n_2 = \infty$, $\sigma_m = 1$)

5.5.2 *The Least-Significant-Difference Test.*

The basic principle of using a preliminary homogeneity of means test to raise a low protection level was first proposed by R. A. Fisher (9). A test which has arisen out of his discussion is the *least-significant-difference test* already mentioned in the introduction.

A general α-level test of this type is given by the rule: *The difference between any two means in a set of n is significant provided that the difference is significant according to an α-level t test and provided also that the variance of the whole set is significant according to an α-level F test.*

In the case of three means, this is identical with an α-level Rule 1 multiple F test with constant levels. The regions of the test under the same special conditions as before are the same as those of the Rule 1 multiple F test with special levels in Figure 5 or the multiple comparisons test in Figure 6 except that the radius of the circle is increased to $\sqrt{4F} = 3.46$, F now being the 5% level value of the F ratio with degrees of freedom 2 and ∞.

In the more general case with n means, $n > 3$, the least significant difference test does not use all of the F tests prescribed by a multiple F test and fails to fix adequate values for all of the protection levels involved. For example in a test of four means, assuming $n_2 = \infty$, $\sigma_m = 1$, $\alpha = .05$ as before, we find $\gamma_2 = 95\%$, $\gamma_3 = 87.8\%$, and $\gamma_4 = 95\%$. The value γ_3 of the three-mean protection levels is as low as that of the corresponding multiple normal deviate test. In general, the value γ_p of any p-mean protection level in an α-level least significant difference test is as low as the γ_p value in the corresponding α-level multiple t test with the one exception that γ_n is raised to $1 - \alpha$.

Thus while this test is more conservative than the new multiple range test or the multiple comparisons test for the case of three means, it is less conservative in cases with more than three means.

5.5.3 *Scheffé's Test Based on Contrasts.*

A recent procedure proposed by Scheffé (19) may be described as the F test analogue of Tukey's test based on allowances.

In the case of three means under the same conditions as before, the regions of this test are generated by the symmetrical intersection of strip regions with straight boundaries like those of the multiple normal-deviate test except that (i) the width of the strips is 2×3.46 instead of 2×2.77, and (ii) the strips are infinite in number as in the Rule 2 multiple F test. The intersections of these strips form a circle of radius 3.46 at the center and this gives the test a three-mean protection level of 95%. At the same time the strip-region protection levels are raised, by the increases in strip-widths, from 95% to 98.6%.

5.6 *Other Decision Procedures.*

As mentioned previously several writers including Bechhofer (1) have dealt with a problem which may be regarded as a special case of the general one with which we have been concerned, and procedures have been proposed which may be regarded as degenerate multiple range or multiple F tests. The decision procedures proposed in the given reference, for example, are for deciding that the t largest means in a sample of n means m_1, m_2, $\cdots$, m_n are all significantly larger than all of the remaining $n - t$ means. In one procedure the true means corresponding to the t largest observed means are not ranked relative to one another; in another procedure they are. In both cases the true means in the remaining subgroup are left unranked relative to one another. To take a simple illustration, in a procedure for choosing the largest mean among four, that is, $t = 1$ and $n = 4$, the decisions

in terms of our previous notation are $(1, 2, 3, 4)$, $(1, 2, 4, 3)$, $(1, 3, 4, 2)$ and $(2, 3, 4, 1)$, where $(1, 2, 3, 4)$, for example, is the decision that μ_4 is larger than each of the remaining means, which are left unranked relative to another.

One very restrictive result of eliminating the missing decisions is that all of the protection levels of the procedure are forced to zero, or in other words all of the significance levels are forced to 100%. For example, in a procedure involving only two means, the experimenter is forced to make the decision $(1, 2)$ or $(2, 1)$. Thus, if it so happens that $\mu_1 = \mu_2$ the probability of making a wrong decision is 100%. The power curves of this test are similar to the $p(1, 2)$ and $p(2, 1)$ curves illustrated for the 5% level test in Figure 1b except that each curve is forced to pass through the 50% power value at $\epsilon = \mu_1 - \mu_2 = 0$. The usefulness of these procedures is therefore restricted to problems in which the experimenter feels impelled to choose a best mean from the results of the given experiment alone.

By limiting themselves to procedures with zero protection levels at the outset, the authors of these tests have been able to avoid the controversial problem of consistent protection levels and to concentrate on other problems such as the tabulation of relations between power functions and sample sizes, (Bechhofer, 1), and the optimum choice of the size of an experiment based on minimax considerations, (Somerville, 20).

6. CONCLUDING REMARKS

Most of the foregoing procedures can be classified usefully according to three basic characteristics:

1. *Type of significant differences*: separating a procedure such as the Newman-Keuls test having a set of significant differences which decrease as the test proceeds, from a procedure such as Tukey's test based on allowances which has one constant significant difference.

2. *Type of protection levels*: separating a procedure such as the Newman-Keuls test having constant values (or lower limits) of $(1 - \alpha)$ for its protection levels*, from a test such as the new multiple range test having protection levels based on degrees of freedom.

3. *Type of component tests*: separating procedures into several categories according to whether they employ range tests, F tests, or component tests of another type.

---

*excluding class 2 protection levels.

Table V shows the allocation of several procedures in a classification of this kind.

The most important of these characteristics is the first, separating tests 1a, with decreasing significant differences, from tests 1b, with constant significant differences. The nature of the confidence interval methods from which the 1b tests are derived is such that in an application of one of these tests there is only one single significant value against which all differences or linear comparisons are tested. This makes for considerable simplicity. However, the single significant value has to be so high that the power functions are severely reduced.

TABLE V. CLASSIFICATION OF TEST PROCEDURES ACCORDING TO THREE BASIC CHARACTERISTICS

2. Type of Protection Levels	1. Type of Significant Differences			
	1a) *Decreasing* 3. Component Tests		1b) *Constant* 3. Component Tests	
	3a) Range	3b) F	3a) Range	3b) F
2a) None less than constant values $\gamma_p = (1 - \alpha)$	Newman-Keuls Test		Tukey's Test Based on Allowances	Scheffé's Test
2b) Protection Levels Based on degrees of freedom $\gamma_p = (1 - \alpha)^{p-1}$	New Multiple Range Test	Multiple Comparisons Test		

For example, in a 5% level Tukey test based on allowances for a case with 20 means (again assuming $n_2 = \infty$, $\sigma_m = 1$), the significant ranges all have the same value 5.01, as shown in Table VI. This value 5.01 is equal to the largest of the significant ranges of the corresponding 1a test, a 5% level Newman-Keuls test, for which the significant ranges, also shown in Table VI, decrease with subset size from 5.01 down to 2.77. In the 1a test, a difference between two means which exceeds only 2.77 can be significant depending on the disposition of the other means. In the 1b test no difference can be significant without exceeding 5.01.

Comparing these two tests further, consider two true means in particular, say μ_1 and μ_2, and suppose that μ_1 is smaller than μ_2. Let

μ_1 and μ_2 on one hand be well separated from the remaining true means μ_3, μ_4, $\cdots$, μ_{20} on the other. For example, suppose $\frac{1}{2}(\mu_1 + \mu_2) = 120$ and $\mu_3 = \mu_4 = \cdots = \mu_{20} = 100$. Under these circumstances, recalling that $\sigma_m = 1$, the observed means m_1 and m_2 will be well separated from the remaining observed means m_3, m_4, $\cdots$, m_{20}. Because of this, the ranges of all subsets of three or more of the observed means which include m_1 and m_2 are practically certain to be significant. Thus in

TABLE VI. COMPARISON OF SIGNIFICANT RANGES FOR 5% LEVEL TESTS OF 20 MEANS

Test	Subset Sizes								
	2	3	4	5	6	8	10	14	20
Tukey's Test Based on Allowances	5.01	5.01	5.01	5.01	5.01	5.01	5.01	5.01	5.01
Tukey's 1953 Test	3.89	4.16	4.32	4.44	4.52	4.65	4.74	4.88	5.01
Newman-Keuls Test	2.77	3.32	3.63	3.86	4.03	4.29	4.47	4.74	5.01
New Multiple Range Test	2.77	2.92	3.02	3.09	3.15	3.23	3.29	3.38	3.47

the 1a test the probability of correctly deciding that μ_1 is less than μ^z will be virtually the same as if the remaining means were not present, that is,

$$p_{1a}(1, 2) = P[\text{dec. } (1, 2) \mid \mu_2 - \mu_1] = P[m_1 - m_2 < -2.77 \mid \mu_2 - \mu_1].$$

For the 1b test, however, the corresponding power is given by

$$p_{1b}(1, 2) = P[\text{dec. } (1, 2) \mid \mu_2 - \mu_1] = P[m_1 - m_2 \leq -5.01 \mid \mu_2 - \mu_1].$$

Table VII shows the values of these two functions and their differences for various values of $\mu_2 - \mu_1$. The differences represent the losses in power in the 1b test relative to the 1a test and some of these can be seen to be very large.

At other parameter values in a 20-mean test, with other arrangements of the true means, the relative losses in power will not be as great. However, it is clear that losses will occur at all values of the parameters and many will be considerable. For tests involving more than 20 means the differences in power will be even greater, increasing as the number of means increases.

TABLE VII.   SEVEREST POWER LOSSES OF 1b TEST RELATIVE TO 1a TEST (5% LEVEL
TESTS OF 20 MEANS)

$\mu_2 - \mu_1$	1a Test	1b Test	Loss
0	.0250	.0002	.0248
1	.1056	.0023	.1033
2	.2946	.0166	.2780
3	.5636	.0778	.4858
4	.8078	.2389	.5689
5	.9429	.4960	.4469
6	.9887	.7580	.2307
7	.9986	.9207	.0779
8	.9999	.9826	.0173
∞	1.0000	1.0000	0.0000

Similar decreases in power must occur in all 1b tests using constant significant differences. These losses appear unnecessary and tests of this type are therefore not recommended.

A partial concession to this point of view is made by Tukey (23) in his 1953 test already mentioned. The significant ranges for this test lie midway between those of the corresponding 1a and 1b tests. An example of these under the conditions already used for the previous 20-mean test examples is also given in Table VI. A test of this type, however, still suffers considerable losses in power probabilities relative to the Newman-Keuls procedure and is also considered to be unnecessarily conservative.

The second most important characteristic is the one concerning protection levels. This separates tests 2a, using constant values (or lower limits) for protection levels, from tests 2b, using the special lower limits based on degrees of freedom.

As has already been mentioned, the power functions of the 2a tests are uniformly lower than those of the corresponding 2b tests. Some further idea of this may be obtained from Table VI by comparing the Newman-Keuls significant ranges, discussed above, with those of the corresponding new multiple range test, which have been taken from Table II, row $n_2 = \infty$.

Each of these tests requires that a difference between any two means must exceed 2.77 before it can be significant and each thus has two-mean protection levels of 95%. The significant ranges for subsets of more than two means, however, are larger in the 2a test. As a result of this, some differences which may not be significant in the 2a test may be significant in the 2b test. It can be seen that the amounts by

which the power functions of the 2b test exceed those of the 2a test are greatest around the origin $\mu_1 = \mu_2 = \cdots = \mu_{20}$ and decrease toward zero in certain directions away from this point. The same holds for any 2b test, relative to the corresponding 2a test.

There appears to be no sound reason for not using protection levels based on degrees of freedom thereby gaining considerably in power to detect real differences.

Finally, there is the subdivision of the test procedures according to the type of component tests employed. In this paper we have considered only procedures based on range tests (3a) and F tests (3b). However, other types of component tests, for example, extreme deviate tests and gap tests, have been proposed and one procedure given by Tukey (21) is based on a combination of three types of component tests.

The problem of deciding the relative merits of various types of component tests is complex, and much work needs to be done in this direction. At present, it appears that the best choice lies between range tests and F tests. The relative merits of these depend on the objectives involved.

Under some circumstances (i), interest may lie in testing linear comparisons involving several means as well as differences between single means; under others (ii), interest may be restricted to testing only differences between single means.

Under circumstances (i) additional power functions are needed to measure the power of the test with respect to the additional comparisons involved. When these are all included it seems safe to assume that multiple F tests are more powerful in some average sense than multiple range tests. Under circumstances (ii), however, the relations are more obscure. The preliminary tests in a multiple F test with decreasing significant differences (1a tests) may cause a little less general interference* with subsequent tests than do the preliminary range tests in a corresponding multiple range test. In this event, the multiple F tests may still be more powerful in an average sense but only slightly so.

The important deciding factor under circumstances (ii) will often be the difference in time and effort required in applying the two types of tests. The application of a multiple range test is much easier and a test of this type will generally be preferred for this reason.

To summarize, the features recommended in each classification are:

1. Decreasing significant differences, as used in tests 1a;

---

*This does not apply of course in 1b tests with *constant* significant differences, in which case the use of range tests gives more powerful procedures. Thus, for example, under circumstances (ii), Tukey's test based on allowances is uniformly more powerful than Scheffé's test

2. Protection levels based on degrees of freedom, as used in tests 2b; and

3. Range tests as used in tests 2a, unless one is interested in linear comparison other than differences between single means, in which case F tests are recommended, as used in tests 3b.

The new multiple range test and the multiple comparisons test have been designed to include these recommended features.

Computation of Tables II and III for New Multiple Range Test.

Let $Q(p, n_2, \alpha)$ represent the entry for given values of p, n_2, and α given in Tables II and III for $\alpha = .05$ and $.01$, respectively. Put $R(p, n_2, \gamma_{p,\alpha})$ for the $100\gamma_{p,\alpha}$ percentage point of the studentized range where $\gamma_{p,\alpha} = (1 - \alpha)^{p-1}$. Then the tabled values have been computed from the relation $Q(p, n_2, \alpha) = R(p, n_2, \gamma_{p,\alpha})$ for $p = 2$, and from $Q(p, n_2, \alpha) = R(p, n_2, \gamma_{p,\alpha})$ or $Q(p - 1, n_2, \alpha)$, whichever is the larger, for all other values of p. This ensures that each p-mean protection level in the new multiple range test is $\gamma_{p,\alpha}$ for all values of p.

The studentized range values $R(p, n_2, \gamma_{p,\alpha})$ for $2 \leq p \leq 20$ and $10 \leq n_2 \leq \infty$ used in this process have been obtained from Pearson and Hartley's Tables (16). The remainder of the $R(p, n_2, \gamma_{p,\alpha})$ values involved have been obtained by new methods (see Beyer, 2) specially developed for this purpose.

Acknowledgment. The author is indebted to W. H. Beyer for much of the theoretical developments and the computational work involved in getting the values $R(p, n_2, \gamma_{p,\alpha})$ of the studentized range required for Tables II and III as explained above.

7. REFERENCES

(1) Bechhofer, Robert E., "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances," *Annals of Mathematical Statistics, 25,* 16–39, 1954.

(2) Beyer, William H., "Certain Percentage Points of the Distribution of the Studentized Range of Large Samples," Virginia Polytechnic Institute M.S. thesis, 56 pp., 1953.

(3) Cornfield, J., Halperin, M. and Greenhouse, S., "Simultaneous Tests of Significance and Simultaneous Confidence Intervals for Comparisons of Many Means," unpublished mimeographed notes, National Institutes of Health, Public Health Service, Department of Health, Education and Welfare, Bethesda, Maryland, 18 pp., 1953.

(4) Davies, Owen L., "Statistical Methods in Research and Production," 2nd ed., London, Oliver and Boyd, 1949.

(5) Duncan, D. B., "Significance Tests for Differences between Ranked Variates Drawn from Normal Populations," Iowa State College Ph.D. thesis, 117 pp., 1947.

(6) Duncan, D. B., "A Significance Test for Differences between Ranked Treatments in an Analysis of Variance," *Virginia Journal of Science*, *2*, 171–189, 1951.

(7) Duncan, D. B., "On the Properties of the Multiple Comparisons Test," *Virginia Journal of Science*, *3*, 49–67, 1952.

(8) Duncan, D. B., "Multiple Range Tests and the Multiple Comparisons Test," (Preliminary Report), *Biometrics*, *9*, Abstract 220, 1953.

(9) Fisher, R. A., "The Design of Experiments," six eds., London, Oliver and Boyd, 1935–1951.

(10) Keuls, M., "The Use of the 'Studentized Range' in Connection with an Analysis of Variance," *Euphytica*, *1*, 112–122, 1952.

(11) Lehmann, E. L., "Some Principles of the Theory of Testing Hypotheses," *Annals of Mathematical Statistics*, *21*, 1–26, 1950.

(12) Newman, D., "The Distribution of the Range in Samples from a Normal Population, Expressed in Terms of an Independent Estimate of Standard Deviation," *Biometrika*, *31*, 20–30, 1939.

(13) Neyman, J., "First Course in Probability and Statistics," Henry Holt and Company, Inc., New York, 1950.

(14) Paterson, D. D., "Statistical Technique in Agricultural Research," McGraw-Hill Book Company, Inc., New York, 1939.

(15) Pearson, E. S., and Hartley, H. O., "The Probability Integral of the Range in Samples of *n* Observations from a Normal Population," *Biometrika*, *32*, 301–310, 1942.

(16) Pearson, E. S., and Hartley, H. O., "Tables of the Probability Integral of the 'Studentized' Range," *Biometrika*, *33*, 89–99, 1943.

(17) Roy, S. N. and Bose, R. C., "Simultaneous Confidence Interval Estimation," *Annals of Mathematical Statistics*, *24*, 513–536, 1953.

(18) Sawkins, D. T., unpublished work, University of Sydney, 1938.

(19) Scheffé, H., "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika*, *40*, 87–104, 1953.

(20) Somerville, P. N., "Some Problems of Optimum Sampling," *Biometrika*, *41*, 420–429, 1954.

(21) Tukey, J. W., "Comparing Individual Means in the Analysis of Variance," *Biometrics*, *5*, 99–114, 1949.

(22) Tukey, J. W., "Quick and Dirty Methods in Statistics," part II, Simple Analyses for Standard Designs, *Proceedings Fifth Annual Convention, American Society for Quality Control*, 189–197, 1951.

(23) Tukey, J. W., "The Problem of Multiple Comparisons," unpublished dittoed notes, Princeton University, 396 pp., 1953.

(24) Wald, A., "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, *10*, 299–326, 1939.

(25) Hartley, H. O., "Some Significance Test Procedures for Multiple Comparisons," *Annals of Mathematical Statistics*, *25*, Abstract 19, 1954.

FURTHER CONTRIBUTIONS TO THE THEORY OF PAIRED COMPARISONS[1]

M. G. Kendall

Visiting Professor, Institute of Statistics, North Carolina State College

1. When a pair of objects is presented for comparison and the two are placed in the relationship preferred: not-preferred, we have what is known as a *paired comparison*. A set of n objects can be compared, a pair at a time, in some or all of the possible $n(n-1)/2$ ways of choosing a pair, and the set of paired comparisons so derived gives us a picture of the interrelationships of the objects under preference. A paired-comparison scheme is more general than a ranking; for with the latter A-preferred-to-B and B-preferred-to-C automatically ensures A-preferred-to-C, whereas with paired comparisons it might happen that C was preferred to A. The existence of these departures from the ranking situation may be due to various reasons, such as the fact that 'preference' is a complicated comparison being made with reference to several factors simultaneously; and one reason for using paired comparisons is to give such effects a chance to show themselves.

2. Situations often occur in which a set of m observers express preferences among n objects and we have to select that object, or perhaps that sub-set of objects, which are, in some sense, "most preferred." The simplest case is the one where there are only two objects, A and B, and every observer votes for either A or B as president of an institution. If 51 per cent of the votes are cast for A and 49 per cent for B we declare A elected. In doing so we have satisfied 51 per cent of the preferences but have had to proceed contrary to 49 per cent; we may say that 49 per cent of the preferences were *violated*. More generally, when we have to select a subset of the n objects as "elected" we shall in general, in the absence of complete unanimity, violate a number of preferences. Circumstances force us to do so to some extent. The problem is to do so to the least possible extent.

3. Consider the case in which 8 members of a body have to elect a committee of three from among themselves. We will suppose that no member votes for himself (though this makes no essential difference) and that there are no abstentions (though this too makes no essential

---

difference). If the 8 members are represented by the letters A to G
they might vote as follows:

Member	Members Preferred
A	BDE
B	DAF
C	DGA
D	CBE
E	ABC
F	ACD
G	BAC

(1)

Here, for the moment, we suppose that there is no preference expressed
among the triplets of members preferred; that is to say, A prefers
B, D, E but does not say whether B is preferred to D or E, or D to E.
He might then have written down his nominees in any order.

Under this system each elector expresses 9 preferences. A, for
example, says, in effect, that he prefers B to C, F, and G, prefers D to
C, F, and G, and prefers E to C, F, and G. There are thus 63 preferences
altogether. We will represent this scheme in a two-way array of the
following kind:

	A	B	C	D	E	F	G	No. of preferences
A	—	11		111	1111	111	111	15
B	1	—	1	11	1	1111	111	12
C	1	1	—	111	111	111	1111	15
D		11	1	—	11	11	11	9
E	1		1		—	11	11	6
F				1	1	—	1	3
G		1			1	1	—	3
Totals	3	6	3	9	12	15	15	63

(2)

Here, if A is preferred to B (a relationship we shall henceforward write
as A pref. B or $A \rightarrow B$) we write a unit in the row A, column B. For
example C prefers D, G, A to each of B, E, F. We therefore have
units in row D, Col. B; row D, Col. E; row D, Col. F; row G, Col. B;

row G, Col. E; row G, Col. F; row A, Col. B; row A, Col. E; row A, Col. F. The totality of preferences expressed in (1) is given in the array (2), together with row and column totals.

Notice that: (a) the sum of row and column totals for each letter is 18. This provides a check. The reason is that each of the letters is compared with three others by each of six observers, so that each letter has 18 preferences (one way or the other).

(b) each column or row total is a multiple of three; for if any letter is preferred at all by an observer it is preferred to three others.

4. From the array (2) we see that A and C had 15 preferences each. If all preferences expressed by all observers have equal weight there is nothing to choose between them. B comes next with 12 preferences. All the others have fewer. Thus, if we have to elect three out of the seven to form a committee, we elect A, B and C.

5. The procedure we have followed exhibits the structure of the preference scheme most clearly, but for the purposes of electing a committee of three we can proceed much more expeditiously. In fact, from array (1) we see that the voting is as follows:

Member	Number of votes
A	5
B	4
C	5
D	3
E	2
F	1
G	1
	—
	21 (3)

A comparison of this with (2) shows that in the latter the row totals are thrice the number of votes. The reason is easy to see, for if any letter gets a vote it is thereby preferred to three others.

6. Now let us suppose that the rules of election are altered slightly and that each elector writes down the three members he prefers *in order of preference*. Such an order might be that of array (1) where, for example, A gives B his first preference, D his second and E his third. Each elector now expresses 12 preferences, three among the set he names and 9 by implication between those three and the three he omits. If we now form an array of preferences we get, instead of (2)

The antisymmetry of the table has now been lost and row or column totals are no longer divisible by three. But we could still pick out the three members with the greatest number of preferences (A, C, B as before) without constructing a full table. In fact from (1) we score for A the following preferences allotted by the electors B to G:

$$4 + 3 + 0 + 5 + 5 + 4 = 21$$

and so for the other letters. The scores are the preference totals in the final column of (4).

	A	B	C	D	E	F	G	Totals
A	—	3	3	4	4	4	3	21
B	2	—	3	3	3	4	3	18
C	2	2	—	4	4	4	4	20
D	1	2	1	—	3	2	3	12
E	1	0	1	0	—	2	2	6
F	0	0	0	1	1	—	1	3
G	1	1	0	0	1	1	—	4
Totals	7	8	8	12	16	17	16	84

$$(4)$$

7. The same method can obviously be applied to any number of voters and any size of committee. Under the condition that there are no abstentions and that nobody votes for himself, the total number of preferences expressed by m voters for a committee of n (no preferences between committee nominees) is $mn(m - n - 1)$; or if preferences are expressed by ranking nominees, is $mn(m - n/2 - \frac{3}{2})$. We may now, if we wish, relax some of the conditions without affecting essentials.

(a) If every man is allowed to vote for himself nothing new is introduced so long as we adhere to the principle of giving each preference the same weight;

(b) The same principles apply when a number of electors express preferences concerning a group of individuals who are not members of themselves. If m judges express preferences for k out of n objects (without ordering them) the number of preferences is $mk(n - k)$.

(c) If there are any abstentions we can continue as before to count those preferences which are expressed. Suppose, for example, that instead of (1) we had the following preferences expressed (second column):

Member	Preferences	Corrected Preferences	
A	BDE	BDE	
B	CA	CA	
C	DGAB	DGA	
D	CBE	CBE	
E	AB	AB	
F	ACD	ACD	
G	BBB	B	(5)

We suppose that these are in order. Member C has overstepped the mark. Unless we reject his ballot as spoiled we delete B from his ordering. Member B prefers C to A and both to the other four, but cannot express a preference between those other four and hence submits only two names. Member G tries to "plump" but we disallow this and count his expression as a preference for B only. We now have the preferences in the third column of (5) giving the following:

Preferences for									
A		4	+3		+5	+5		= 17	
B	5			+4	+4		+5	= 18	
C		5		+5		+4		= 14	
D	4		+5			+3		= 12	
E	3			+3				= 6	
F								= 0	
G			4					= 4	
								71	(6)

A, B and C are still elected but B now gains more preferences than A.

We notice that election on this principle maximizes the number of satisfied preferences as before.

(d) If any voter "ties" certain nominees, this is equivalent to expressing no preference between them and everything proceeds as before. For example, if in (5) member D tied C, B, E there would be two fewer preferences for C and one fewer preference for B in (6).

(e) In particular this method covers the case when each of a set of judges ranks all the objects, and not merely a preferred sub-set of them. The whole method, in fact, is very flexible in this respect. So long as

any preferences are expressed we can pursue the same technique. The only thing to take particular care about is that one judge has the same *opportunities* as another for expressing the same number of preferences, even though he may not avail himself of them. We clearly introduce bias if we give one judge a chance to express two preferences and another only one. The system proposed is in accordance with the best democratic principles in that each judge has the same number of votes, and all votes have the same weight.

(f) It is possible to order the members, according to the number of preferences allotted to them, in a ranking (which may itself contain tied members). Thus we constrain a paired-comparison system into a ranking at the expense of violating a number of preferences. The fewer the violations the nearer the scheme to an actual ranking. In tables of the type of (2) or (4) a perfect and unanimous ranking would correspond to a situation in which all the non-zero cells were above the main diagonal.

(g) In those cases where we choose to regard any object as compared with itself, as for example if we wish to complete the diagonals in (2), we may allot $\frac{1}{2}$ to the cell in the same row and column. This will clearly not affect the order of the objects according to numbers of preferences received, for each object then receives an extra $\frac{1}{2}$ for each observer.

(h) Likewise, if an observer cannot express a preference between a given pair A, B we may allot $\frac{1}{2}$ to each of the cells in row A, column B and row B, column A in arrays of type (2).

(i) We can, if we wish, give effect to differences in reliability between judges. For example, if in array (2) we regard D as twice as important in his preferences as the others, we enter 2 for each preference instead of unity in the table.

8. Finally, let us note that the number of preferences can be used to calculate a coefficient of agreement among judges. This is another aspect of the coefficient of agreement in paired comparisons proposed by Babington Smith and myself some years ago. (See my *Advanced Theory of Statistics*, vol. 1, chapter 16). In fact if the total possible number of agreements is N and the actual number of agreements is M, the coefficient of agreement would be simply $2M/N - 1$ which varies from $-1/m$ or $-1/(m-1)$ to 1. In table (2) for example the cells (A, B) and (B, A) have respectively 2 and 1 members. The pair A, B are compared three times and of these comparisons two are in agreement; there is thus one agreement out of a possible 3; likewise for AG, there are three agreements, each in the all AG, out of a possible 3. For the whole table it will be found that there are 47 agreements out of a possible 74 and the coefficient of agreement is 0.270.

9. We may also use the table to calculate a coefficient of departure from the ranking situation. Suppose we arrange the table so that rows and columns follow the order of the number of preferences expressed; in the case of table (2) this merely amounts to interchanging the rows and columns corresponding to B and C. The number of units below the diagonal is then 13 and that above the diagonal is 50. No other arrangement of rows and columns can divide the 63 preferences so unequally. If all were above the diagonal the preferences would be consistent with a ranking. We might then take as our measurement of departure from the ranking situation the coefficient $(13/63) \times 2 = 0.413$. We have multiplied the factor 13/63 by two because the furthest situation from ranking occurs when *one half* of the total preferences are allotted to the cells below the diagonal.

10. So much for the elements of the subject. I now proceed to consider sundry developments which are necessary to enable a more penetrating study of a paired-comparison situation to be made. The first arises from the nature of paired comparisons in themselves and may best be introduced by an example.

Let us suppose that six players A to F are engaged in a chess tournament in which each plays the other once. The set of scores (1 for a win, $\frac{1}{2}$ for a draw and 0 for a loss) then represents a set of paired comparisons made in all possible ways between them. We assume that all games reach a decision so that there are no missing values. A possible set of results is as follows:

	A	B	C	D	E	F	Total score
A	$\frac{1}{2}$	1	1	0	1	1	$4\frac{1}{2}$
B	0	$\frac{1}{2}$	0	1	1	0	$2\frac{1}{2}$
C	0	1	$\frac{1}{2}$	1	1	1	$4\frac{1}{2}$
D	1	0	0	$\frac{1}{2}$	0	0	$1\frac{1}{2}$
E	0	0	0	1	$\frac{1}{2}$	1	$2\frac{1}{2}$
F	0	1	0	1	0	$\frac{1}{2}$	$2\frac{1}{2}$

(7)

The simple way of arranging the competitors in order of success is to add up their scores, as is done in the final column. If we had three prizes we should divide the first and second between A and C and divide the third among B, E and F. Only D does not qualify for a share of the prize money. Such a procedure would be adopted in most tournaments of the kind.

11. But we now notice one rather anomalous effect. D, the only

player to receive nothing, has in fact beaten one of the winners, A. We are not allowed to dismiss this as a mere fluke, because all preferences are equally valid. Furthermore A has beaten C but is nevertheless ranked with him. Vague but genuine feelings for general equity lead us to inquire whether something should not and cannot be done to restore the balance. Such a method was suggested by Dr. T. H. Wei (1952) in an unpublished thesis successfully submitted to the University of Cambridge for the Ph.D. degree. In effect Wei's procedure amounts to this:

We recalculate a score for each player by giving him the score of every player he has beaten and half the score of every player with whom he has drawn. This leads to the following new scores:

$$A = \tfrac{1}{2}(4\tfrac{1}{2}) + \; 2\tfrac{1}{2} \; + \; 4\tfrac{1}{2} \; + \; 0 \;\; + \; 2\tfrac{1}{2} \; + \; 2\tfrac{1}{2} \; = \; 14\tfrac{1}{4}$$

$$B = \; 0 \;\; + \tfrac{1}{2}(2\tfrac{1}{2}) + \; 0 \;\; + \; 1\tfrac{1}{2} \; + \; 2\tfrac{1}{2} \; + \; 0 \;\; = \; 5\tfrac{1}{4}$$

$$C = \; 0 \;\; + \; 2\tfrac{1}{2} \; + \tfrac{1}{2}(4\tfrac{1}{2}) + \; 1\tfrac{1}{2} \; + \; 2\tfrac{1}{2} \; + \; 2\tfrac{1}{2} \; = \; 11\tfrac{1}{4}$$

$$D = \; 4\tfrac{1}{2} \; + \; 0 \;\; + \; 0 \;\; + \tfrac{1}{2}(1\tfrac{1}{2}) + \; 0 \;\; + \; 0 \;\; = \; 5\tfrac{1}{4}$$

$$E = \; 0 \;\; + \; 0 \;\; + \; 0 \;\; + \; 1\tfrac{1}{2} \; + \tfrac{1}{2}(2\tfrac{1}{2}) + \; 2\tfrac{1}{2} \; = \; 5\tfrac{1}{4}$$

$$F = \; 0 \;\; + \; 2\tfrac{1}{2} \; + \; 0 \;\; + \; 1\tfrac{1}{2} \; + \; 0 \;\; + \tfrac{1}{2}(2\tfrac{1}{2}) = \; 5\tfrac{1}{4}(8)$$

We now arrange the players in order of new scores; and we now notice that A and C have separated, A being first and C second, while D has moved up to equality with B, E, and F.

This is as far as one would wish to go on practical grounds, perhaps, but now a further point raises itself. We have re-allocated the scores once. Why not do so again? If we re-allocate the scores of (8) by the same method we find

$$A \quad \tfrac{1}{2}(14\tfrac{1}{4}) + \; 5\tfrac{1}{4} \; + 11\tfrac{1}{4} + 0 \; + 5\tfrac{1}{4} + 5\tfrac{1}{4} = 34\tfrac{1}{8}$$

$$B \quad\;\; 0 \;\; + \tfrac{1}{2}(5\tfrac{1}{4}) + \; 0 \; + 5\tfrac{1}{4} + 5\tfrac{1}{4} + 0 \; = 13\tfrac{1}{8}$$

$$C \qquad\qquad\qquad\qquad\qquad\qquad\qquad\;\; = 26\tfrac{5}{8}$$

$$D \qquad\qquad\qquad\qquad\qquad\qquad\qquad\;\; = 16\tfrac{7}{8}$$

$$E \qquad\qquad\qquad\qquad\qquad\qquad\qquad\;\; = 13\tfrac{1}{8}$$

$$F \qquad\qquad\qquad\qquad\qquad\qquad\qquad\;\; = 13\tfrac{1}{8} \qquad (9)$$

A and C are still first and second but D takes third place and B, E, F share the fourth position.

If we re-allocate the scores once more we find scores

A	824.375
B	365.625
C	695.625
D	425.625
E	365.625
F	365.625

(10)

The order is now the same as we derived from (9); and if we ascertain new scores on the same principle we shall find that no new ordering has appeared. Later I shall prove that after a time the situation always "settles down" in this way.

13. There are two interesting features of this procedure. Let us revert to the preference scheme of (7) and regard the scores as a matrix. If we square this matrix we obtain

$$
\begin{bmatrix}
\frac{1}{4} & 3 & 1 & 4 & 2 & 3 \\
1 & \frac{1}{4} & 0 & 2 & 1 & 1 \\
1 & 2 & \frac{1}{4} & 4 & 2 & 2 \\
1 & 1 & 1 & \frac{1}{4} & 1 & 1 \\
1 & 1 & 0 & 2 & \frac{1}{4} & 1 \\
1 & 1 & 0 & 2 & 1 & \frac{1}{4}
\end{bmatrix}
\qquad
\begin{array}{c}
Row\ totals \\
14\frac{1}{4} \\
5\frac{1}{4} \\
11\frac{1}{4} \\
5\frac{1}{4} \\
5\frac{1}{4} \\
5\frac{1}{4}
\end{array}
$$

(11)

and the row totals are those previously obtained in (8) by the first re-allocation of scores. The reason for this will be obvious to anyone familiar with the rules of matrix multiplication and the result is generally true for all preference matrices. Furthermore, if we multiply (11) again by the matrix (7) and add row totals we shall get the scores of (9); and so on. The continual re-allocation of scores is equivalent to taking successive powers of the matrix.

14. Let us now consider what interpretation can be given to the process in terms of comparisons. The following diagram shows the scheme of (7) in geometrical form. The six players are represented by the six vertices of a regular hexagon, which are joined by straight lines in all possible ways. If A pref. B we draw an arrow from A towards B. If no preference was expressed (or the game was drawn) we do not draw an arrow.

15. It will be seen that the score of any player in (7) is the number of arrows *leaving* his vertex, together with $\frac{1}{2}$ (as the conventional score in the diagonal, when he is compared with himself) and $\frac{1}{2}$ for any line passing through his vertex on which no arrow is drawn. When we proceed to the next stage we count the number of paths leaving the
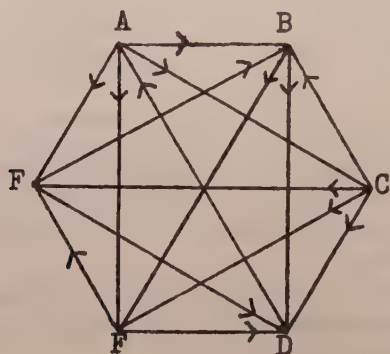
(12)

FIGURE 1

vertex and taking two steps. For example, for A we have the following paths leaving A and also leaving the vertex next visited:

$ABD, ABE; \quad ACB, ACD, ACE, ACF; \quad AED, AEF; \quad AFB, AFD.$

There are ten of these "transitive" preferences. We also count the preference of B with itself, C with itself, etc., as $\frac{1}{2}$ each, making a further score of 2; and finally we score $\frac{1}{2}$ of $\frac{1}{2}$ for the double preference of A with itself. The total score is $14\frac{1}{4}$, which is the score for A in (11). It may be verified that the same procedure gives the other scores in that array.

Similarly the scores obtained by the next re-allocation, as given in (9), are the numbers of paths of three lines leaving the respective vertices, all arrows going the same way, with similar conventions about vertices taken with themselves; and so on. Our re-allocation is equivalent to powering the matrix or to counting paths of transitive preferences of increasing extent.

16. From the geometrical viewpoint it is seen that in proceeding by re-allocation we are extending our concept of comparison. We began by considering comparisons of pairs by themselves. When we proceed to the next stage we compare pairs which form part of triads; but we do not compare the triads by considering them as three pairs (which would bring us back to the first situation). Thus it is possible to

"compare" A and C by the route $A \to B \to C$ or A and B by $A \to C \to B$. Both of these "comparisons" do not count in our score because they cannot both happen together; but either counts when it occurs.

17. Or we may put it another way by saying that we compare two members AB not directly, but through their comparisons with other members, e.g. by ACB, ADB, AEB and AFB. We choose the leading members in the final order so as to maximize the agreement with transitive preferences. Whether this is the right thing to do depends to some extent on practical circumstances. The process of continual re-allocation has the advantage that it results in an objective final ordering; but whether this is what we want depends on whether we are considering a situation in which direct comparison is the basic generator of the data, or whether we wish to give scope for more reflective judgment in roundabout comparisons involving other members.

18. Let us now consider the case when several judges make paired comparisons, or several tournaments are played between the same set of players. For each observer we shall have a preference matrix of the type of (7). To obtain a composite picture, on the supposition that the judges are equally reliable, we superpose the matrices. Thus if (7) represents the preferences of a judge for 6 varieties of ice cream when offered to him in pairs, two additional judges might have the following preference matrices:

	A	B	C	D	E	F	Totals
A	$\frac{1}{2}$	1	0	1	1	0	$3\frac{1}{2}$
B	0	$\frac{1}{2}$	$\frac{1}{2}$	1	0	1	3
C	1	$\frac{1}{2}$	$\frac{1}{2}$	1	0	1	4
D	0	0	0	$\frac{1}{2}$	1	$\frac{1}{2}$	2
E	0	1	1	0	$\frac{1}{2}$	1	$3\frac{1}{2}$
F	1	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	2	(13)

	A	B	C	D	E	F	Totals
A	$\frac{1}{2}$	0	1	1	$\frac{1}{2}$	1	4
B	1	$\frac{1}{2}$	1	0	0	1	$3\frac{1}{2}$
C	0	0	$\frac{1}{2}$	1	1	1	$3\frac{1}{2}$
D	0	1	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$2\frac{1}{2}$
E	$\frac{1}{2}$	1	0	$\frac{1}{2}$	$\frac{1}{2}$	1	$3\frac{1}{2}$
F	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	1	(14)

Adding these and (7) together we get

	A	B	C	D	E	F	Totals
A	$1\frac{1}{2}$	2	2	2	$2\frac{1}{2}$	2	12
B	1	$1\frac{1}{2}$	$1\frac{1}{2}$	2	1	2	9
C	1	$1\frac{1}{2}$	$1\frac{1}{2}$	3	2	3	12
D	1	1	0	$1\frac{1}{2}$	$1\frac{1}{2}$	1	6
E	$\frac{1}{2}$	2	1	$1\frac{1}{2}$	$1\frac{1}{2}$	3	$9\frac{1}{2}$
F	1	1	0	2	0	$1\frac{1}{2}$	$5\frac{1}{2}$
Totals	6	9	6	12	$8\frac{1}{2}$	$12\frac{1}{2}$	54

(15)

On the basis of simple paired comparisons we should place A and C as bracketed equal, E as third, B as fourth, D as fifth and F as last.

19. The question now arises whether we should re-allocate the scores by powering the matrix (15); or whether it would be preferable to power each matrix and then amalgamate the rankings at the end. The two processes will not always lead to identical results, although in practice they should not differ very much. Arithmetically it is simpler to power just the one matrix (15), and in cases where there are many judges this would be almost decisive. This is the procedure I would recommend myself, but if there were any serious doubts I would perform the analysis both ways and compare the results. A wide disparity would, in my view, suggest that neither was very reliable. It would arise mostly in cases where there were substantial disagreements among judges.

20. I now prove that the process of repeated powering does in fact converge to a limiting ranking. Dr. Wei offered a proof of the result for one observer and a complete set of preferences in his thesis.

First of all we define a matrix A of non-negative elements to be *indivisible* if it cannot be expressed in the form (by rearrangement of rows and columns)

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \tag{16}$$

If a preference matrix of type (15) is divisible in this sense the members of one block of objects are always preferred to every member of another. In such a case we divide the data into the two blocks and operate on each, finally ranking the members of the first group and then the members of the second. Similarly, if one of these blocks is itself divisible we divide it up; and so on. We clearly lose no generality by doing this, and divisibility is not a handicap in our preference situations.

21. I now require a theorem of Frobenius (cf. Wielandt, 1950[2]) which says that for indivisible matrices A with non-negative elements and positive elements in the diagonal there exists a unique simple positive root of the equation $|A - \lambda I| = 0$ which is greater than all other roots in absolute value; and that the corresponding characteristic vector has all its elements of the same sign (which we may take to be positive).

Let λ_1 be this largest root and Y_1 the corresponding vector. Then if $\lambda_2, \cdots \lambda_p$ are the other roots and $Y_2 \cdots Y_p$ the corresponding vectors, and if P be the preference matrix, we have

$$PY = \Lambda Y \tag{17}$$

where Λ is the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{bmatrix} \tag{18}$$

It is now easy to show that for any positive integer k

$$P^k Y = \Lambda^k Y \tag{19}$$

As the powering proceeds the major root λ_1 becomes dominant and (19) tends to the equation

$$P^k Y_1 = \lambda_1^k Y_1 \tag{20}$$

Thus from some k onwards the final ordering will be determined by the vector Y_1, which has non-negative elements.

22. We notice that the proof remains applicable to preference matrices in which some preferences may be missing, or when ties are present, provided that the matrix is not divisible. If any cell in a combined preference matrix contains no entries we insert a zero.

23. It is also of some interest to note that we may prove that the preference matrix is never singular. In fact, we can always express it (apart from positive numerical factors) in the form

$$(Q + U) \tag{21}$$

---

where Q is an anti-symmetric matrix and U is the matrix all of whose elements are unity. For example (15), after division of rows by $1\frac{1}{2}$, can be expressed as U plus the matrix

$$Q = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & 0 & 0 & 1 & \frac{1}{3} & 1 \\ -\frac{1}{3} & -\frac{1}{3} & -1 & 0 & 0 & -\frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & -\frac{1}{3} & 0 & 0 & 1 \\ -\frac{1}{3} & -\frac{1}{3} & -1 & \frac{1}{3} & -1 & 0 \end{bmatrix} \tag{22}$$

We reduce $Q + U$ systematically by subtracting the first column from the second column, then the first row from the second row; then the first column from the third column, then the first row from the third row; and so on. The effect on Q is to reduce it to another antisymmetric matrix, say Q', and the effect on U is to reduce it to a unit in the top left-hand corner and zero elsewhere. Thus the determinant of $Q + U$ is the determinant of Q' plus the determinant of the principal minor obtained by omitting the first row and column, which is also antisymmetric.

Now the determinant of $p \times p$ antisymmetric matrix is zero if p is odd and positive if p is even. Hence the determinant of $Q + U$ is the sum of two components, one zero and the other positive; and hence it does not vanish.

22. In practice the number of paired comparisons arising from n objects may be inconveniently large and the question arises whether it is possible to economize in the number of comparisons made. In the example of the chess tournament which has been mentioned above (paragraph 10) if each player is to play every other, 15 games must be played. But only three can be conducted at once, so at best 5 sessions are necessary. If this is too long, and, say, three sessions are all that can be allowed, only nine games can be played and six have to be sacrificed. The question is, which six? Or again, if an individual is comparing items by taste testing, his patience or his palate may not endure the presentation of all the possible pairs, and a problem arises as to how best to cut down the number of pairs and which pairs to present.

23. Problems like this arise in many fields of experimentation and are usually dealt with by incomplete balanced blocks. Some new points, however, arise in paired-comparison work. Durbin (1951) has considered

the use of Youden designs in ranking experiments. More recently Benard and van Elteren (1953) have discussed tests of significance where incomplete rankings are concerned. Without trying to exhaust the subject I proceed to consider the use of incomplete balanced blocks in preference schemes.

24. Consider first of all the case of a single observer. Of the $n(n - 1)/2$ preferences which he could make we require to pick out a sub-set. Certain elementary principles of choice at once suggest themselves:

(a) every object should appear equally often. In this sense the design should be balanced;

(b) the preferences should not be divisible in the sense that we can split the objects into two sets and no comparison is made between any object in one and any object in the other.

In terms of preference matrices (a) means that there should be the same number of non-empty items in each row and column; (b) means that the matrix does not divide into two blocks and become of the form $\left(\begin{smallmatrix} X & 0 \\ 0 & Y \end{smallmatrix}\right)$ when the zeros represent empty cells. In terms of the preference diagram (a) means that there are the same number of paths direct between points leaving or entering each vertex and (b) means that the figure does not separate into two distinct polygons.

25. When possible I add a further condition of symmetry to the situation, that is to say

(c) In the preference diagram the number of paths of length l proceeding from any point to any other point shall be the same for all pairs of points.

The length l here means the number of lines traversed in the path, e.g. the path (in Figure 1, section 14) ABC from A to C is of length 2 and $AEBDC$ from A to C is of length 4. Where no pair of objects is compared in these "partial" situations we omit the line between them. If they are joined by a line without an arrow this means that they have been compared but that no preference has been expressed.

In terms of preference matrices this condition implies a kind of symmetry of interlocking. A path ABC implies entries in row A, column B and column C (and the reflections column A, row B and row C); and analogous entries must occur in other rows in such a way that all the objects are symmetrically involved.

26. Under these conditions we can meet a requirement suggested— to me in conversation by Dr. R. C. Bose: if all the preferences are exerted at random (e.g. if we toss up for it which of a pair shall be preferred) all possible final orderings of the objects produced by powering the matrix should be equally probable. This follows from the symmetry

of the situation, for we can interchange two objects in the designs without altering the preference matrix, so far as concerns the underlying probabilities, and all final orders are therefore equally probable.

27. In a sense, it seems to me, condition (c) is necessary as well as sufficient for a proper design. If it is not obeyed certain objects become subject to different schemes of preference from others and their final positions are not determined on an unbiased basis. In terms of powered preference matrices, the sums of rows are not based on the same number of transitive comparisons of length l.

28. The conditions laid down above impose certain restrictions on the scope of a paired-comparison experiment. For instance, if there are six objects and the numbers of entries in the rows of the preference matrix are equal, the number of comparisons necessary to obtain a balanced experiment must be a multiple of three. Anything else destroys the balance. The connectivity condition (b) further limits the freedom of choice; for example, with six objects at least six comparisons are required.

29. The setting up of incomplete designs is most easily thought of in terms of tours round the preference polygon. Consider the case $n = 7$. (Prime numbers are easier to deal with in most experimental designs.) There are 21 comparisons altogether. To obtain a balanced design we must have either 7 or 14 comparisons (or, of course, the full 21). The first 7 may, without loss of generality, be taken as the tour $ABC \cdots G$ round the preference heptagon. (No generality is lost
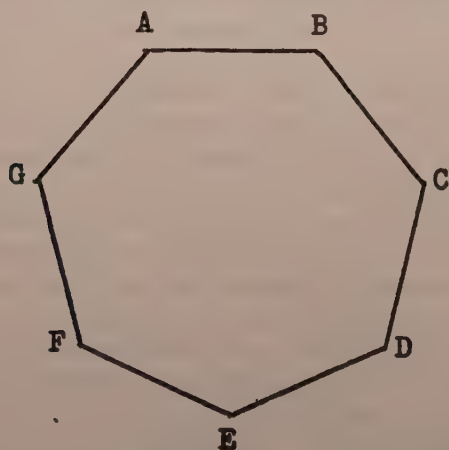

FIGURE 2

because each member must be connected to two others and hence they are on a chain which may be taken to be the order A to G.) For

the next 7 we have two possibilities: (a) start from A, miss a vertex and go to C, miss a vertex and go to E and so on; (b) start from A, miss two vertices and go to E, then two vertices and go to G and so on. We do not obtain new designs by tours missing three or more vertices because they are equivalent to (a) or (b). The two schemes are shown in Figure 3.
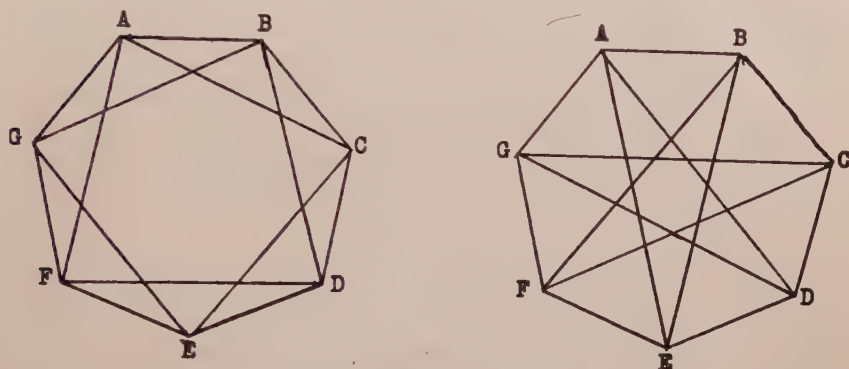

FIGURE 3

These schemes are not identical. In the former there are two triangular tours connecting any pair, e.g. ACB and AGB, whereas in the second there is only one, e.g. AEB. In terms of time taken in perform-ance there is nothing to choose between them. For example if they represented a chess tournament, each round requires three games, one player having a bye, and for 14 games 5 rounds are required. Such might be

Scheme 1			Bye	Scheme 2			Bye	
AB,	CD,	EF	G	AB,	CD,	EF,	G	
AC,	BD,	EG	F	AD,	BC,	FG,	E	
BC,	DE,	FG	A	BE,	CF,	DG,	A	
AF,	CE,	BG	D	AE,	BF,	CG,	D	
AG,	DF		B, C, E	AG,	DE		B, C, F	(23)

30. It remains to be considered whether one scheme is preferable to the other by some other criterion. There is nothing to choose between them in relation to balance or the application of the powered-matrix method. We note, however, that the patterns of transitive preferences are different. In the first any pair is connected by two triangles, three

quadrilaterals, etc., in the second by one triangle, four quadrilaterals, etc. On the whole, I should be inclined to select the second design from a feeling that it has higher connectivity, but an exact criterion awaits further investigation.

31. When we have several judges, an obvious extension of symmetry requirements necessitates that each participates to an equivalent extent: in some sense the design should be balanced by judges as well as by comparisons. Something depends on whether we require to compare judges in addition to objects. If so, each pair of judges must have certain comparisons in common. With two judges and seven objects, for example, one simple way would be to allot to each 14 comparisons, one judging according to each of the designs of Figure 3. They would then have 7 comparisons in common and all possible comparisons could be made.

32. I do not propose on this occasion to attempt a systematic exposition of the design problems involved in paired comparisons. Designs of an optimum kind which balance by numbers of comparisons, objects compared, numbers of observers on given comparisons and so forth are probably rather rare; and if something has to be sacrificed it depends on what is the point of major interest whether we sacrifice symmetry in comparisons or in judges. A final example will make clear a few of the principles involved.

Consider again the case of seven objects, $ABCDEFG$. There are three distinct tours round the preference polygon,

$$
\begin{array}{ccccccc}
A & B & C & D & E & F & G \\
A & C & E & G & B & D & F \\
A & D & G & C & F & B & E
\end{array} \tag{24}
$$

Each tour involves seven comparisons and each object is compared with two others in a tour.

For a complete set of comparisons each observer would have to make 21. If this is felt to be too much we may allocate 14, consisting of two tours each. And if the tours are represented by a, b, c, we may allocate to the observers 1, 2, 3

$$
\begin{array}{lll}
1: & a, & b \\
2: & b, & c \\
3: & c, & a
\end{array} \tag{25}
$$

With these schemes every comparison is made equally often (twice); every tour is made equally often (twice); every observer makes the

same number of comparisons (14); every observer has a tour in common with every other observer; and thus every observer can be compared with every other observer in respect of two comparisons involving any specified object.

If we have more than three observers, we take a number equal to a multiple of three and replicate the design.

Now suppose we had eleven objects, A to K. The full set of comparisons numbers 55. There are five distinct tours round the preference polygon

$$
\begin{array}{llllllllllll}
a: & A & B & C & D & E & F & G & H & I & J & K \\[4pt]
b: & A & C & E & G & I & K & B & D & F & H & J \\[4pt]
c: & A & D & G & J & B & E & H & K & C & F & I & (26) \\[4pt]
d: & A & E & I & B & F & J & C & G & K & D & H \\[4pt]
e: & A & F & K & E & J & D & I & C & H & B & G
\end{array}
$$

Now if we try to allot two tours to each of five observers we lose symmetry; for there are 10 pairs of tours choosable from these five. We have, to preserve complete balance, to allot four tours to each observer 1, 2, 3, 4, 5

$$
\begin{array}{lllll}
1: & b, & c, & d, & e \\[4pt]
1: & c, & d, & e, & a \\[4pt]
3: & d, & e, & a, & b \\[4pt]
4: & e, & a, & b, & c \\[4pt]
5: & a, & b, & c, & d & (27)
\end{array}
$$

Again the tours are balanced, but we have not achieved very much. Each observer now makes 44 comparisons, against the full set of 55.

We can sacrifice symmetry in several ways. We may, for instance, allot two tours to each observer, e.g.

$$
\begin{array}{lll}
1: & a, & b \\[4pt]
2: & b, & c \\[4pt]
3: & c, & d \\[4pt]
4: & d, & e \\[4pt]
5: & e, & a & (28)
\end{array}
$$

Here every observer can be compared with two other observers but not every pair can be compared. Or if we have, say, 10 observers we may allot all the 10 possible pairs of tours one to each. Each observer then makes 22 comparisons and can be compared with four other observers. If 22 comparisons are still felt to be too many for one observer we may allocate the 55 preferences according to a linked design, e.g. (numbering the preferences 1 to 55) with 11 observers, 10 preferences each

1 :	1,	2,	3,	4,	5,	6,	7,	8,	9,	10	
2 :	1,	11,	12,	13,	14,	15.	16,	17,	18,	19	
3 :	2,	11,	20,	21,	22,	23,	24,	25,	26,	27	
4 :	3,	12,	20,	28,	29,	30,	31,	32,	33,	34	
5 :	4,	13,	21,	28,	35,	36,	37,	38,	39,	40	
6 :	5,	14,	22,	29,	35,	41,	42,	43,	44,	45	
7 :	6,	15,	23,	30,	36,	41,	46,	47,	48,	49	
8 :	7,	16,	24,	31,	37,	42,	46,	50,	51,	52	
9 :	8,	17,	25,	32,	38,	43,	47,	50,	53,	54	
10 :	9,	18,	26,	33,	39,	44,	48,	51,	53,	55	
11 :	10,	19,	27,	34,	40,	45,	49,	52,	54,	55	(29)

Here we have cut down the comparisons for each observer to 10 and each comparison is made twice. But we have lost a good deal of the comparison between judges; every judge can be compared with every other judge but only on one comparison of objects.

REFERENCES

Benard, A. and Van Elteren, Ph. (1953), A generalization of the method of m rankings. *Kon. Neder. Ak. van Weterschappen.* A, 56, 358.

Durbin, J. (1951), Incomplete blocks in ranking experiments. *Brit. Jour. Psych.* 4, 85.

Frobenius, G. (1912), Über Matrizen aus nicht negativen Elementen. *Sitz. Preuss. Akad. Wiss.*, 456.

Wei, T. H. (1952), The algebraic foundations of ranking theory. Unpublished thesis, Cambridge University, England.

Wielandt, H. (1950), Unzerlegbare, nicht negative Matrizen. *Math. Zeit.*, 52, 642.

COMPARATIVE SENSITIVITY OF PAIR AND TRIAD FLAVOR INTENSITY DIFFERENCE TESTS[1]

J. W. Hopkins and N. T. Gridgeman

Division of Applied Biology, National Research Council, Ottawa, Canada

INTRODUCTION

Alternative simple experimental designs for sensory difference tests of flavor intensity lead to the procedures termed "pair", "duo-trio" and "triangular" tests (3). In the first, a unit trial consists in submitting coded aliquots of the two batches in question to a subject in the sequence (X), (Y) or (Y), (X) and requiring him to rank them in the order of appraised flavor strength. In the second, it consists in submitting identified X or Y with the coded sequence (X), (Y) or (Y), (X) and requiring the subject to attempt to match the identified with the like coded aliquot. In the third, it consists in submitting one of the completely coded sequences (X), (X), (Y); (X), (Y), (X); (Y), (X), (X); (Y), (Y), (X); (Y), (X), (Y) or (X), (Y), (Y) and again requiring an attempted matching of like aliquots. Inferences respecting the occurrence or non-occurrence of real discrimination are then made by relating the actual frequency of ranking or matching in repeated trials to percentage points of the binomial distribution expected in the absence of discrimination.

It has been suggested (4) that the "triangular" test is "obviously the most efficient" but experimental evidence to the contrary has been reported (1). This note indicates some statistical considerations relevant to efficiency comparisons, and applies them to additional data.

STATISTICAL CONSIDERATIONS

At a nominal significance level of 5% for a "pair" test n-replicated, the appropriate critical region for rejection of the null hypothesis that sensorily $X = Y$ will comprise all x at or below the effective 2.5% and at or above the effective 97.5% points of the cumulative binomial distribution (6, 7) of x for n and $p_0 = 1/2$. Here p_0 is the chance probability on the null hypothesis and x the recorded frequency of a specified ranking, e.g. of (Y) above (X). For the "duo-trio" and "triangular" tests, which involve only matching without ranking, the corresponding critical regions will comprise all points x at or above the effective 95% points of the cumulative binomial distribution of x for n and $p_0 = 1/2$ and $p_0 = 1/3$ respectively. Here p_0 is the probability
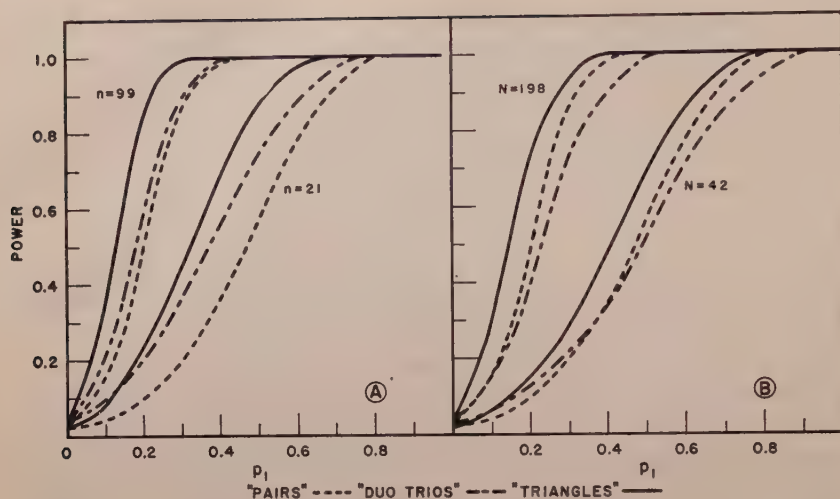

FIG. 1

Power of "pair", "duo-trio" and "triangular" flavor intensity appraisals for detection of $Y > X$ when nominal significance level $\alpha = 0.05$ in relation to the probability p_1 of genuine sensory discrimination: (A) for equal numbers of replicates $n = 21$ and $n = 99$; (B) for equal numbers of aliquots $N = 42$ and $N = 198$.

on the null hypothesis and x the recorded frequency of matching like aliquots.

In the presence of a marginal difference having a constant probability p_1 of sensory recognition, the probability p of ranking (Y) above (X) in a "pair" test will be the sum of p_1 and of the conditional probability of chance guessing after failure to discriminate. Hence $p = p_1 + (1 - p_1)/2$ or $p = 1 - p_1 - (1 - p_1)/2$, i.e. $p = (1 \pm p_1)/2$, according as the intensity of $X > Y$ or of $Y < X$. For "duo-trio" and "triangular" tests the probability of matching like aliquots will now correspondingly be $p = (1 + p_1)/2$ and $p = (1 + 2p_1)/3$ respectively. Fig. 1A depicts the resulting power at nominal $\alpha = 0.05$ of these three tests of $X \neq Y$, i.e. the probability $1 - \beta$ that x will fall in the critical regions specified, as p_1 ranges from 0 to 1 when $n = 21$ and when $n = 99$. For corresponding $0 \leq p_1 \leq 1$, the power order is "pair" < "duo-trio" < "triangular". However for equal numbers N of appraised aliquots (Fig. 1B) the order is "duo-trio" $\gtreqless$ "pair" < "triangular", the power of "pair" relative to "duo-trio" tests varying with N and p_1 partly because of differences between the nominal and effective percentage points of discrete distributions. Unfortunately, in practice p_1 is unspecifiable *a priori*. An assumption of equal p_1 in all three types of test for an identical flavor contrast or the consequences of inequalities in p_1 must be tested experimentally in specific instances.

Three parallel trials were made in the writers' laboratory. Slight modifications of the flavor intensity of an aqueous solution generating a mixture of the four primary tastes (Trial A), of tomato juice (Trial B) and of minced steak (Trial C) provided three flavor contrasts of differing complexity. These were each appraised under standardized conditions by six experienced subjects in 18 replicate "pair", "duo-trio" and "triangular" discrimination tests, of which there were thus 972 in all. Sequences of presentation of the various coded aliquot pairs and triads to each subject were randomized subject to the condition that each of the two possible coded pair sequences occurred equally frequently, likewise each of the four "duo-trio" sequences X, (X), (Y); X, (Y), (X); Y, (X), (Y) and Y, (Y), (X), and likewise each of the six triad sequences enumerated above.

Table I summarizes the results obtained. In all nine tests the recorded total frequency of specified rankings or matchings exceeded its no-discrimination expectation of 54 for the "pair" and "duo-trio" and of 36 for the "triangular" tests.

TABLE I.

Recorded Frequency of Specified Ranking and Matching of Aliquots in (1) "Pair", (2) "Duo-Trio" and (3) "Triangular" Flavor Tests

Trial and test	Subject						Total x
	I	II	III	IV	V	VI	
A.1	9	15	9	14	13	15	75
A.2	14	11	9	8	13	10	65
A.3	5	8	12	8	8	9	50
B.1	10	11	9	10	14	7	61
B.2	7	9	10	12	11	9	58
B.3	5	5	9	6	6	7	38
C.1	12	12	14	13	12	9	72
C.2	10	9	10	11	13	13	66
C.3	6	8	12	10	8	9	53

ANALYSIS OF DATA

Intra-test homogeneity

Calculated indices of dispersion (Cochran's (2) Q), appropriate to repetitive data for the same individuals (5), were entirely consistant with inter-replicate stability of sensory discrimination by the group of

subjects as a whole. Moreover, the individual frequencies of specified rankings and matchings listed in each row of Table I, when arrayed together with their complements in nine 2×6 contingency tables, gave an aggregate index of inter-subject intra-test homogeneity of $\chi^2 = 45.9$ with $9 \times 5 = 45$ d.f. In this instance therefore it is also reasonable to assume that p was sensibly the same for all six subjects in any one test and trial.

Inter-test differences

The logarithm of the likelihood of any recorded x for "pair" and "duo-trio" tests will be:

$$\log L = x \log \left(\frac{1 + p_1}{2}\right) + (n - x) \log \left(\frac{1 - p_1}{2}\right),$$

while for "triangular" tests

$$\log L = x \log \left(\frac{1 + 2p_1}{3}\right) + (n - x) \log \left(\frac{2 - 2p_1}{3}\right).$$

Hence maximum likelihood estimates $\hat{p}_1$ of $p_1 > 0$, specified by equating $\partial \log L / \partial p_1$ to zero, will result from $(2x - n)/n$ for "pair" and "duo-trio" and from $(3x - n)/2n$ for "triangular" tests. "Pair" and "duo-trio" tests in which $x \leq n/2$ and "triangular" tests in which $x \leq n/3$ provide no internal evidence of $p_1 > 0$. As $\hat{p}_1$ is a linear function of $\hat{p} = x/n$, the random sampling variance of the former will be $V(\hat{p}_1) = 4V(\hat{p})$ for "pair" and "duo-trio" and $9V(\hat{p})/4$ for "triangular" tests; and with $n = 108$, $V(\hat{p})$ may be estimated with reasonable confidence from $\hat{p}(1 - \hat{p})$. From the marginal totals of Table I, the following $\hat{p}_1$ result.

Trial	Test		
	"Pair"	"Duo-trio"	"Triangular"
A	.39	.20	.19
B	.13	.07	.03
C	.33	.22	.24
Average	.28	.16	.15

The difference of 0.125 between the mean $\hat{p}_1$ for the "pair" and that for both the 3-aliquot tests is 1.97 times its estimated standard deviation of 0.0635. The mean $\hat{p}_1$ for the "duo-trio" and "triangular" tests evidently do not differ significantly.
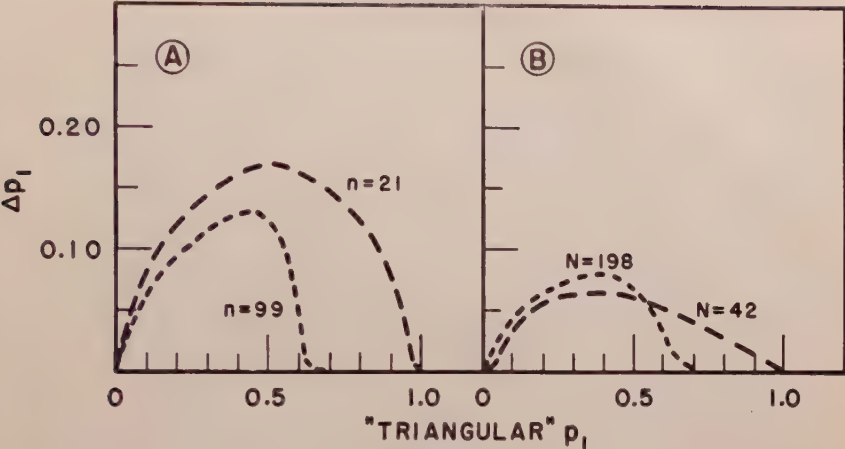
FIG. 2

Increment Δp_1 in the probability p_1 of genuine sensory discrimination required to equalize the power of "pair" and "triangular" flavor intensity appraisals, in relation to "triangular" p_1 : (A) for equal numbers of replicates $n = 21$ and $n = 99$; (B) for equal numbers of aliquots $N = 42$ and $N = 198$.

Power Effects

Discriminatory powers for $X \neq Y$ specifically attained in these experiments cannot be estimated with exactitude, because of the imprecision in $\hat{p} = x/n$ with n no larger than 108. However, since all three trials were consistent with identical p_1 for "duo-trio" and "triangular" appraisals, the relative sensitivity of these may be inferred from Fig. 1 and the preceding $\hat{p}_1$. For p_1 the same as the listed $\hat{p}_1$ the comparative powers $1 - \beta$ for detecting $Y > X$ with equal numbers n of replicates and N of aliquots, and nominal significance level $\alpha = 0.05$, would be:

Trial	"Pair" $n = 21$ $N = 42$	"Triangular" $n = 21$ $N = 63$	$n = 14$ $N = 42$	"Pair" $n = 99$ $N = 198$	"Triangular" $n = 99$ $N = 297$	$n = 66$ $N = 198$
A	.40	.29	.13	1.00	.79	.67
B	.08	.06	.02	0.16	.08	.09
C	.30	.38	.19	1.00	.92	.82

Fig. 2 illustrates the increment Δp_1 required for equipotency of "pair" and "triangular" tests as a function of "triangular" p_1 in the equal replicate and equal aliquot instances exemplified above. Ordinates of the curves depicted in this figure correspond to abscissal distances between power curves in Fig. 1.

DISCUSSION

For corresponding p_1, "triangular" tests have a statistical advantage over "duo-trios" and "pairs", both per replicate and per aliquot. However, the preceding experimental results, together with those of Byer and Abrams (1), suggest that in some instances at least p_1 may in fact be greater in "pair" appraisals, possibly because fewer inter-comparisons are required. The data also suggest that such discriminatory superiority may sometimes more than offset the statistical advantage per aliquot of "triangles".

Man-hours devoted to flavor appraisals are not all spent in actual tasting. Appreciable aggregate amounts of time may also be required to schedule, assemble, instruct and return subjects to their own working quarters. These are largely independent of the number of aliquots appraised per session. When the latter is small therefore the relative power per man-hour of "pair" and "triangular" tests may be intermediate between their relative powers per replicate and per aliquot. In large-scale testing, and whenever test materials are scarce or costly, relative power: cost ratios will approximate more closely to the latter. Also, matching tests may be applicable to the detection of qualitative differences for which ranking is inappropriate. Factors such as these, as well as purely statistical considerations, may accordingly influence a rational choice between pair and triad tests for specific applications.

REFERENCES

1. Byer, A. J. and Abrams, D.   A comparison of the triangular and two-sample taste-test methods. *Food Technology, 7*: 185–187, 1953.
2. Cochran, W. G. The comparison of percentages in matched samples. *Biometrika,* 37: 256–266, 1950.
3. Dawson, E. H. and Harris, B. L. Sensory methods for measuring differences in food quality. *Agric. Information Bull. 34,* U.S. Dept. of Agriculture, Washington, D. C., 1951.
4. Harrison, S. and Elder, L. W. Some applications of statistics to laboratory taste-testing. *Food Technology, 4*: 434–439, 1950.
5. Hopkins, J. W. Some observations on sensitivity and repeatability of triad taste difference tests. *Biometrics, 10*: 521–531, 1954.
6. National Bureau of Standards. Applied Mathematics Series 6, Tables of the Binomial Probability Distribution. U.S. Gov't. Printing Office, Washington, D. C., 1950.
7. Romig, H. G. *50-100 Binomial Tables.* John Wiley and Sons, Inc., New York, 1953.

THE DESCRIPTION OF GENIC INTERACTIONS IN CONTINUOUS VARIATION

B. I. HAYMAN AND K. MATHER

*A.R.C.'s Unit of Biometrical Genetics,
Department of Genetics, University of Birmingham*

The genetical interpretation of the continuous variation (or indeed any variation) shown by a population, family or group of families requires the use of specifications of two distinct kinds. Firstly, it is necessary to specify the genetical structure of the population, family or families. In principle, this requires the specification in suitable terms of the relative frequencies of the various alleles of the genes involved, the distribution of the alleles at a locus between the various possible homozygotes and heterozygotes and the distribution of the alleles of different genes in respect of one another. These specifications will depend on the ancestry of the material, the mating system which has been in force, the selection which has been practised (if any), and the linkage or other relation of the genes in transmission from parent to offspring. Specification of the genotype of every individual, or indeed of any individual, is not essential for most biometrical purposes so long as the relative frequencies of the different possible genotypes can be given, and indeed it is sufficient for many purposes to specify only the average, taken over all genes, of the allele frequencies, homozygosis, linkage relations and so on.

Secondly, it is necessary to specify the relations between genotype and phenotype. In principle, this requires specification of the effect of each gene substitution on the character or characters in question, the dominance relations of the genes, the relations in effect of non-allelic genes (genic interaction), the effects of non-heritable agencies, and the relations in effect of genic and non-heritable agencies (genotype-environment interaction). Specification of the effects of heritable but extra-nuclear particles may also be required in special cases, but experience shows that these may generally be neglected. Like the genetical specifications, the specifications of effect need not, for most biometrical purposes, be individually detailed. It will suffice for many purposes to specify only the effects of gene substitution, dominance and genic interaction, each pooled over all genes, and the pooled effects of all non-heritable agencies and their interactions, so that neither the individual genes need be isolated nor the non-heritable agencies separated.

Given these two sets of specifications, the phenotypic properties (which are the properties capable of being observed) of the population can be predicted prior to observation; individually for each member of the population if the specifications are individually detailed, or statistically for the whole population if, as is usually the case, the specifications are statistical. We are, however, more commonly concerned with deriving the specifications from the observed properties of the material. This is clearly impossible without some knowledge of the genetical relations or breeding behaviour of the individuals whose phenotypes are observed. Generally it has been found convenient to set up the experiments in such a way that certain genetical specifications can be reasonably assumed. Thus if we start with a cross between two true-breeding strains of plants and proceed thereafter by self-pollination, all precautions being taken to avoid selection, we may assume that only two alleles are present at any locus and that the rise of homozygosis will follow the rule Mendel first enunciated, both within and between the lineages derived at various stages of the experiment. Linkage relations remain to be inferred from the observations themselves, and if the possibility of selection has not been eliminated, it may be no easy matter to distinguish between the effects of linkage and selection (Bateman and Mather 1951). Other systems of mating—sib-mating, backcrossing, diallel crossing and so on—may be, and have been, used for the same purpose; and inversions may be used so to reduce recombination within chromosomes that the linkage relations are simplified to an extent where they may be reasonably assumed. The basic principle of the approach remains the same in all these cases, and it depends for its success on the demonstration that all but a negligible fraction of the heritable component of continuous variation springs from nuclear genes, whose behaviour in transmission is understood from other types of genetical investigation. In this way the genetical study of continuous variation rests on the foundation provided by mendelian genetics in all its complexity and strength. Where determination is extra-nuclear, the genetical specification alters, and becomes less certain. It may indeed then cease to be a matter for confident assumption and become one for investigation and inference.

The specification of effect is seldom if ever capable of the same precise assumption, for the reason that no generalisations of a precision and breadth of application comparable to those of the chromosome theory of heredity are available in respect of gene action. True, we are guided to the extent that we must bargain for genic interactions, both allelic in the form of dominance and non-allelic in the form of epistasis, and also genotype-environment interactions. In this way

we are told the broad classes into which the variation, or rather its causation, must be partitioned; but we do not know in detail what to expect. Many types of interaction may exist side by side and we have no means of anticipating any one type or any mixture of types. Specification of effect is thus one of our regular and prime tasks of inference in interpreting continuous variation. One general tool we do have, however. The specification of effect will vary with the scale on which the character is measured. It is therefore assumed that this scale has been chosen to minimise the various types of interaction. Tests are available of the validity of this assumption (Mather 1949a). There can be no certainty, however, that a scale exists on which all interactions will vanish, and indeed we have evidence in particular cases that while scaling may reduce, it cannot wholly remove, the interactions that are present. Any comprehensive consideration of the specification of effect must therefore take into account interactions between non-allelic genes. In attempting such consideration our first task is clearly that of arriving at a suitable way of describing and classifying the interactions. It is with this first aspect of the problem that we are concerned in the present account.

The Description of Interactions

In diploid organisms the individual can fall into any one of three genetic classes (AA, Aa and aa) in respect of a gene for which there exist two alleles (A and a).* Two independent comparisons are possible among three classes. The effect of the gene difference on the phenotype can thus be described completely by two parameters, and specified completely if the values of these two parameters are known. Statistically the pair of parameters may be defined in a variety of ways, but these will not all be of equal value in genetical analysis. In the system adopted by Fisher *et al.* (1932) and Mather (1949 a and b), one parameter (d) is used to represent the phenotypic difference between the two homozygotes, AA and aa, and the other (h) to represent the departure in phenotype of the heterozygote, Aa, from the mid-point between AA and aa. Taking this mid-point as the origin, the effects on the phenotype are then

$$
\begin{array}{ccc}
aa & Aa & AA \\
-d & h & d
\end{array}
$$

so that the gene's contribution to the fixable genetic variation is pro-

---

*A is used to denote the allele tending to increase the manifestation of the character, and not, as is conventionally the case, to denote the dominant allele. The direction of dominance is indicated by the sign of the parameter h.

portional to d, while h reflects the dominance properties of the gene and represents the contribution to the unfixable heritable variation. At the same time, the contributions of d and h to the heritable variation will be statistically independent so long as the two homozygotes are equally frequent in the population or families. When this condition is not fulfilled, their contributions to the variation will be partly confounded.

With two gene differences, nine genotypes* are possible, and eight

TABLE 1

	$\begin{array}{c}\textbf{AA}\\ d_a\end{array}$	$\begin{array}{c}\textbf{Aa}\\ h_a\end{array}$	$\begin{array}{c}\textbf{aa}\\ -d_a\end{array}$
$\begin{array}{c}\textbf{BB}\\ d_b\end{array}$	$d_a + d_b$ $+ i_{ab\|}$ $- \frac{1}{2}j_{a\|b} - \frac{1}{2}j_{b\|a}$ $+ \frac{1}{4}l_{\|ab}$	$h_a + d_b$ $+ \frac{1}{2}j_{b\|a}$ $- \frac{1}{4}l_{\|ab}$	$-d_a + d_b$ $- i_{ab\|}$ $+ \frac{1}{2}j_{a\|b} - \frac{1}{2}j_{b\|a}$ $+ \frac{1}{4}l_{\|ab}$
$\begin{array}{c}\textbf{Bb}\\ h_b\end{array}$	$d_a + h_b$ $+ \frac{1}{2}j_{a\|b}$ $- \frac{1}{4}l_{\|ab}$	$h_a + h_b$ $+ \frac{1}{4}l_{\|ab}$	$-d_a + h_b$ $- \frac{1}{2}j_{a\|b}$ $- \frac{1}{4}l_{\|ab}$
$\begin{array}{c}\textbf{bb}\\ -d_b\end{array}$	$d_a - d_b$ $- i_{ab\|}$ $- \frac{1}{2}j_{a\|b} + \frac{1}{2}j_{b\|a}$ $+ \frac{1}{4}l_{\|ab}$	$h_a - d_b$ $- \frac{1}{2}j_{b\|a}$ $- \frac{1}{4}l_{\|ab}$	$-d_a - d_b$ $+ i_{ab\|}$ $+ \frac{1}{2}j_{a\|b} + \frac{1}{2}j_{b\|a}$ $+ \frac{1}{4}l_{\|ab}$

The phenotypes associated with the nine genotypes in respect of two interacting genes.

parameters must be used to give a complete description of the phenotypes. Four of these will be the d's and h's appropriate to the two genes, as shown in the margins of Table 1. The other four may then be derived conveniently to correspond to the "interaction" comparisons

---

*With two linked genes, there are ten genotypes because the double heterozygotes fall into the two classes AB/ab and Ab/aB. Generally, however, these genotypes give a common phenotype so that the distinction of linkage phase need be pursued no further in our present discussion.

of an analysis of variance where the d's and h's correspond to the "main effects". The distribution of these four parameters among the nine genotypes are shown in Table 1. They fall into three classes. One of these, $i_{ab|}$ is the interaction of d_a and d_b and may be termed the homozygote-homozygote interaction. Two others, $j_{a|b}$ and $j_{b|a}$ are the homozygote-heterozygote interactions, respectively, of d_a and h_b, and d_b and h_a. the last, $l_{|ab}$, is the heterozygote-heterozygote interaction of h_a and h_b. The coefficients of $\frac{1}{2}$ and $\frac{1}{4}$ are applied to the j's and l respectively so that equal contributions will be made to the overall differences in an F_2 family by interactions of unit size. The double frequency of heterozygotes in an F_2 also makes it unnecessary to vary the coefficients of j and l from cell to cell of the table.

The four interactions, as defined in this way, have clear genetical meanings, though they do not follow the conventional genetical classification of interaction between non-allelic genes. All the classical types of interaction may, however, be cast in terms of i, j and l. The standard mendelian F_2 segregation into four phenotypic classes with frequencies 9:3:3:1 occurs when $d_a = h_a$, $d_b = h_b$ and $i_{ab|} = j_{a|b} = j_{b|a} = l_{|ab}$. Thus although this type of F_2 is classically regarded as showing no interaction of the genes, interactions may be present within certain restrictions. If we add the further condition that $d_a = \frac{3}{2} i_{ab|}$ we obtain the 9:3:4 ratio characteristic of recessive epistasis. The further condition that $d_a = d_b$, then gives the 9:7 ratio of complementary genes.

Going back to the standard F_2, the additional condition $d_a = -\frac{1}{2}i_{ab|}$ gives the 12:3:1 ratio of dominant epistasis; while the addition of the still further condition $d_a = d_b$ gives the 15:1 ratio of duplicate factors. Again, if instead of this last condition we put $d_a = -\frac{1}{3}d_b$, the 13:3 ratio of the recessive suppressor relation is obtained. Indeed any interaction of two genes can be achieved by imposing appropriate conditions. For example, a situation which might be described as dominant dominance modification results from putting $d_a = 2h_a = 2d_b = 2h_b = j_{b|a} = l_{|ab}$ and $i_{ab|} = j_{a|b} = 0$. These various relations are shown diagrammatically in Table 2.

The representations of all the classical types of interaction in terms of the same parameters, i, j and l, permit their combination in analysis, so that it becomes possible to consider any number of genes with many diverse interactions between them without any further elaboration. Each will contribute in its own way to the i, j and l components of variation and so long as we can discover how these components change from generation to generation we can give an average, or statistical, account of the interactions and their effects on variation without having to aim at any individual classification. Furthermore, as we shall see

TABLE 2

The classical F_2 and six types of classical digenic interaction in terms of d, h, i, j and l.

CLASSICAL F_2

$$d_a = h_a \quad d_b = h_b$$
$$i = j_{a|b} = j_{b|a} = l$$

DOM. DOM. MODIFIER

$$d_a = 2h_a = 2d_b = 2h_b = j_{b|a} = l$$
$$i = j_{a|b} = 0$$

	AA	**Aa**	**aa**
BB	$d_a + d_b$ $+\frac{1}{4}i$		$-d_a$ $+d_b$
Bb			$-\frac{3}{4}i$
bb	$d_a - d_b - \frac{3}{4}i$		$-d_a - d_b$ $+\frac{9}{4}i$

$\frac{5}{4}d_a$ $-\frac{3}{4}d_a$

DOM. EPISTASIS

$$d_a = -\tfrac{1}{2}i$$

$\frac{1}{2}d_a + d_b$

$\frac{5}{2}d_a - d_b$ $-\frac{1}{2}\tfrac{1}{}d_a$ $-d_b$

REC. EPISTASIS

$$d_a = \tfrac{3}{2}i$$

$\frac{7}{6}d_a + d_b$ $-\frac{3}{2}d_a$ $+d_b$

$\frac{1}{2}d_a - d_b$

DUPLICATE GENES

$$d_a = d_b$$

$\frac{3}{2}d$

$-\frac{1}{2}\tfrac{3}{}d$

REC. SUPPRESSOR

$$d_a = -\tfrac{1}{3}d_b$$

$-\frac{5}{2}d$

$\frac{1}{2}\tfrac{1}{}d$

COMPLEMENTARY GENES

$$d_a = d_b$$

$\frac{1}{6}\tfrac{3}{}d$

$-\frac{1}{2}d$
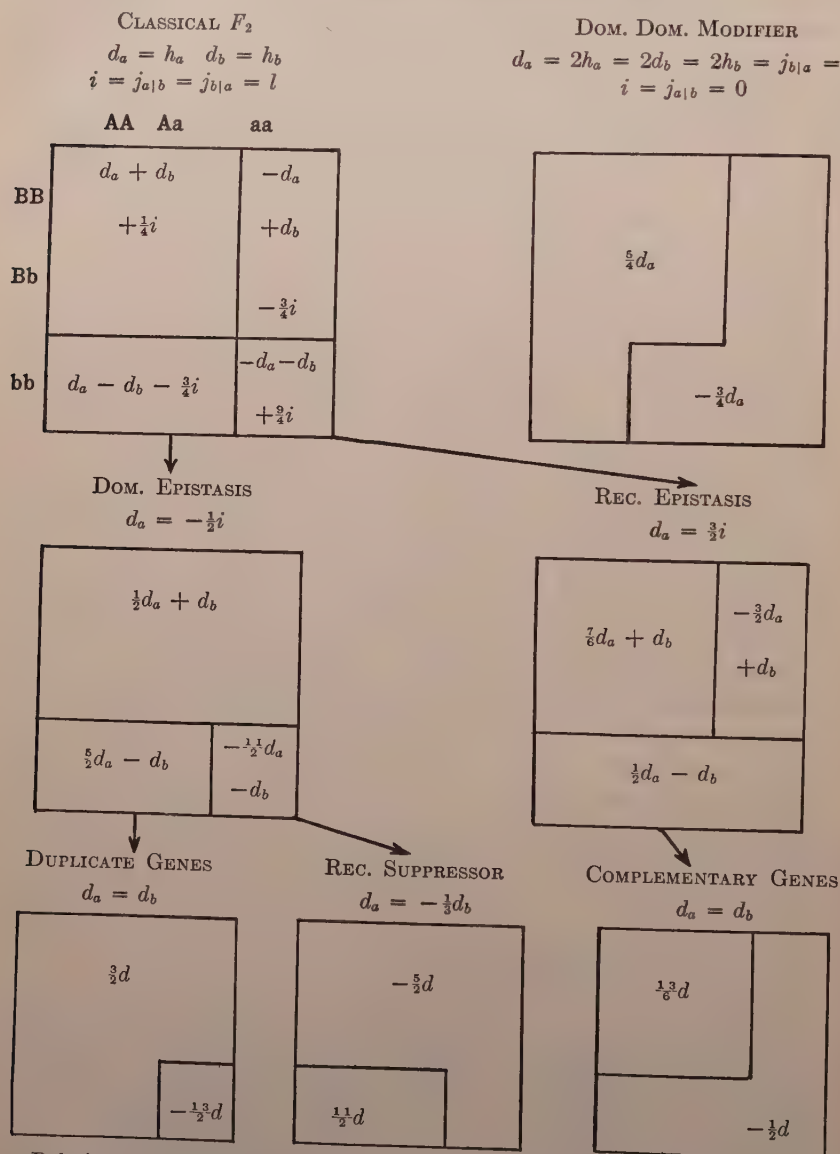
Relations among those parameters which yield the Classical F_2 (top left) and the Dominant Dominance Modifier (top right) interaction are shown in full. The remaining classical interactions are derived from the Classical F_2 by the addition of further relations between the parameters as shown above each square. The class phenotypes are shown within the squares.

below, the three categories i, j and l, have their own properties of effect and change with the generations, so that the classification into these catagories is, in principle, sufficient to enable us to understand, estimate and predict, the effects of interactions between pairs of genes.

This system of classification can be extended to interactions between three or more genes. With trigenic interactions we should recognise four categories, hom-hom-hom, hom-hom-het, hom-het-het and het-het-het. So four new types of parameter would come in to the analysis, though two of the types would each include three individual parameters, making eight in all. To describe the phenotypes of the 27 genotypes produced by three genes requires 26 parameters. Of these 18 are already available, 6 from the two parameters describing the main effects of each of three genes, and 12 from the four parameters describing the digenic interactions among the three pairs possible with three genes. The 8 parameters required for the trigenic interactions complete the tally.

The phenotype is found as the algebraic sum of all the parameters associated with the genotype in question (Table 1). So the sum of the "main effect" parameters, d and h, gives a first approximation to the phenotype—one which neglects all interactions. Thus for the genotype $AABB$ in Table 1, we should have $d_a + d_b$ as this first description of the phenotype. Moving to the next level of approximation by admitting digenic interactions (which in this simple two gene model is the final approximation giving a complete description) we define the phenotype as $d_a + d_b + i_{ab1} - \frac{1}{2}(j_{a|b} + j_{b|a}) + \frac{1}{4}l_{|ab}$. With a polygenic model we can obtain a next approximation by bringing in the eight parameters for trigenic interactions and so on. These successive approximations might, however, be expected to become of less and less advantage. Most of the variation will generally (though not, of course, necessarily) be accounted for by the "main effect" parameters, most of the rest by the parameters for digenic interactions and so on. There will thus be little justification for considering the more complex interactions until the digenic type has been fully explored.

The Effects of Interactions

The contributions of the two interacting genes to the mean expression of the character in the various generations derivable from a cross between two true-breeding lines, are shown in Table 3. All increments are measured from the mid-parent value, which is of course the mean of the expression in the two parental lines. Two crosses are possible, in respect of the two genes: one where the increasing allelomorphs of the two genes are associated in one parent, and the decreasing allelo-

TABLE 3

Generation Means in Respect of Two Interacting Genes

Parents: Associated $\begin{matrix}\bar{P}_1\\\bar{P}_2\end{matrix}$ $\pm d_a \pm d_b + .i_{ab|} \mp \tfrac{1}{2}j_{a|b} \mp \tfrac{1}{2}j_{b|a} + \tfrac{1}{4}l_{|ab}$

Dispersed $\begin{matrix}\bar{P}_1\\\bar{P}_2\end{matrix}$ $\pm d_a \mp d_b - i_{ab|} \mp \tfrac{1}{2}j_{a|b} \pm \tfrac{1}{2}j_{b|a} + \tfrac{1}{4}l_{|ab}$

Backcrosses: Associated $\begin{matrix}\bar{B}_1\\\bar{B}_2\end{matrix}$ $\tfrac{1}{2}(\pm d_a \pm d_b + h_a + h_b + \tfrac{1}{2}i_{ab|})$

Dispersed $\begin{matrix}\bar{B}_1\\\bar{B}_2\end{matrix}$ $\tfrac{1}{2}(\pm d_a \mp d_b + h_a + h_b - \tfrac{1}{2}i_{ab|})$

$\bar{F}_1 \quad h_a + h_b + \tfrac{1}{4}l_{|ab}$

$\bar{F}_2 \quad \tfrac{1}{2}(h_a + h_b)$

$\bar{F}_3 \quad \tfrac{1}{4}(h_a + h_b + \tfrac{1}{4}l_{|ab})$

$\bar{S}_3 \quad \tfrac{1}{2}(h_a + h_b)$

SCALING TESTS

		Associated	Dispersed
A	$\bar{P}_1 + \bar{F}_1 - 2\bar{B}_1$	$\tfrac{1}{2}(i_{ab|} - j_{a|b} - j_{b|a} + l_{|ab})$	$\tfrac{1}{2}(-i_{ab|} - j_{a|b} + j_{b|a} + l_{|ab})$
B	$\bar{P}_2 + \bar{F}_1 - 2\bar{B}_2$	$\tfrac{1}{2}(i_{ab|} + j_{a|b} + j_{b|a} + l_{|ab})$	$\tfrac{1}{2}(-i_{ab|} + j_{a|b} - j_{b|a} + l_{|ab})$
C	$\bar{P}_1 + \bar{P}_2 + 2\bar{F}_1 - 4\bar{F}_2$	$2i_{ab|} + l_{|ab}$	$-2i_{ab|} + l_{|ab}$
D	$\bar{P}_1 + \bar{P}_2 + 2\bar{F}_2 - 4\bar{F}_3$	$2i_{ab|} + \tfrac{1}{4}l_{|ab}$	$-2i_{ab|} + \tfrac{1}{4}l_{|ab}$

"Associated" refers to the cross in which increasing and decreasing allelomorphs of the two genes occur in the same parents (AABB × aabb); and "Dispersed" to the alternative cross (AAbb × aaBB). Where two signs are shown before a term, the upper and lower signs used in the formulae refer respectively to the upper and lower families shown on the left.

morphs in the other ($AABB \times aabb$); and another where each parental line carries the increasing allelomorph of one gene and the decreasing allelomorph of the other ($AAbb \times aaBB$). These are referred to respectively as "Associated" and "Dispersed" distributions of the genes. The mean expressions in the parental families ($\bar{P}_1$ and $\bar{P}_2$) and in the families raised by backcrossing the F_1 to the parents ($\bar{B}_1$ from $F_1 \times P_1$, and $\bar{B}_2$ from $F_1 \times P_2$) vary with genic distribution in the parents; but the means of F_1, F_2, F_3 and the biparental third generation or S_3 (raised by random crossing among the individuals of F_2) are independent of distribution in the absence of linkage. Free recombination of the genes is assumed in all these formulae, and indeed in the whole of the present discussion.

The values to be expected from the scaling tests (Mather 1949a) are shown at the bottom of Table 3. These all reduce to zero when no interaction is present, but each type of test depends on characteristic sets of interactions for its departure from zero. In other words, each type of scaling test is capable of detecting its own characteristic constellation of interactions. Where the mean of F_3 is available, D provides a test largely of the i type interaction. Test C depends to a greater extent on the l type interaction, and so provides a means of assessing both i and l interactions when used in conjunction with D. The j type interactions have no effect on tests C and D, but will affect the outcome of the backcross tests, A and B. Combinations of these tests can obviously be devised to detect specific types of interaction.

It should be observed that with a particular distribution of genes between the parents A and B may afford only insensitive tests of j interactions, for these may in part cancel out. The sign of the effect of the i interaction in all tests also varies with the distribution of the genes in the parent lines, but the contribution made by the l interaction is unaffected by genic distribution. Furthermore, where more than two interacting genes are affecting the variation of the character in a cross, so that there may be two or more interactions of each kind, the different i and l interactions, as well as the j interactions, may tend to balance one another's effects if the directions of the individual i's and l's vary. Or to put it in other words, if, for example, of $i_{ab|}$, $i_{ac|}$, $i_{bc|}$ etc. some are acting in the $+$ direction and others in the $-$ direction, the sum of these i's (which will appear in the scaling tests) may well be low because of the balancing relations of the different i's one against another.

This balancing action, introduced by differences in sign, is always likely to be encountered in the contributions made to means and comparisons between them. It is less troublesome when we turn to the effect of interactions on the second degree statistics which we calculate from segregating generations. The contribution of two interacting genes to the variances and covariances obtained from backcrosses, F_2, F_3 and S_3 are given in Table 4. The variance of F_2 (V_{1F2}) includes separate items for each type of interaction, and since these items are all quadratic, the contribution will be unaffected by sign. The same is true of the three statistics, (V_{1S3}, V_{2S3} and W_{1S23}) obtainable from the S_3 generation.

The situation in the case of the F_3 statistics is more ambiguous. A portion of the effect of each type of interaction still appears as a separate quadratic item, and indeed the whole of the effect of the i interaction is expressed in this way. The j and l interactions, on the other hand, become partly confounded with the d and h items respectively, in the

<div align="center">

TABLE 4

Variances and Covariances in Respect of Two Interacting Genes

</div>

Summed variances of backcrosses:

$$V_{B1} + V_{B2} : \text{Associated} = \tfrac{1}{2}\{(d_a - \tfrac{1}{2}j_{b|a})^2 + (d_b - \tfrac{1}{2}j_{a|b})^2 + (h_a - \tfrac{1}{2}i_{ab|})^2$$
$$+ (h_b - \tfrac{1}{2}i_{ab|})^2 + \tfrac{1}{4}(i_{ab|} + l_{|ab})^2 + \tfrac{1}{4}(j_{a|b} + j_{b|a})^2\}$$
$$\text{Dispersed} = \tfrac{1}{2}\{(d_a + \tfrac{1}{2}j_{b|a})^2 + (d_b + \tfrac{1}{2}j_{a|b})^2 + (h_a + \tfrac{1}{2}i_{ab|})^2$$
$$+ (h_b + \tfrac{1}{2}i_{ab|})^2 + \tfrac{1}{4}(i_{ab|} - l_{|ab})^2 + \tfrac{1}{4}(j_{a|b} - j_{b|a})^2\}$$

Variance of F_2,

$$V_{1F2} = \tfrac{1}{2}d_a^2 + \tfrac{1}{2}d_b^2 + \tfrac{1}{4}h_a^2 + \tfrac{1}{4}h_b^2 + \tfrac{1}{4}i_{ab|}^2 + \tfrac{1}{8}j_{a|b}^2 + \tfrac{1}{8}j_{b|a}^2 + \tfrac{1}{16}l_{|ab}^2$$

Variance of F_3 means,

$$V_{1F3} = \tfrac{1}{2}(d_a - \tfrac{1}{4}j_{a|b})^2 + \tfrac{1}{2}(d_b - \tfrac{1}{4}j_{b|a})^2 + \tfrac{1}{16}(h_a - \tfrac{1}{4}l_{|ab})^2 + \tfrac{1}{16}(h_b - \tfrac{1}{4}l_{|ab})^2$$
$$+ \tfrac{1}{4}i_{ab|}^2 + \tfrac{3}{32}j_{a|b}^2 + \tfrac{3}{32}j_{b|a}^2 + \tfrac{1}{256}l_{|ab}^2$$

Mean variance of F_3 families,

$$V_{2F3} = \tfrac{1}{4}(d_a - \tfrac{1}{4}j_{a|b})^2 + \tfrac{1}{4}(d_b - \tfrac{1}{4}j_{b|a})^2 + \tfrac{1}{8}(h_a - \tfrac{1}{4}l_{|ab})^2 + \tfrac{1}{8}(h_b - \tfrac{1}{4}l_{|ab})^2$$
$$+ \tfrac{5}{16}i_{ab|}^2 + \tfrac{7}{64}j_{a|b}^2 + \tfrac{7}{64}j_{b|a}^2 + \tfrac{3}{32}l_{|ab}^2$$

Covariance of F_2 and F_3 family means,

$$W_{1F23} = \tfrac{1}{2}d_a(d_a - \tfrac{1}{4}j_{a|b}) + \tfrac{1}{2}d_b(d_b - \tfrac{1}{4}j_{b|a}) + \tfrac{1}{8}h_a(h_a - \tfrac{1}{4}l_{|ab}) + \tfrac{1}{8}h_b(h_b - \tfrac{1}{4}l_{|ab})$$
$$+ \tfrac{1}{4}i_{ab|}^2 + \tfrac{1}{16}j_{a|b}^2 + \tfrac{1}{16}j_{b|a}^2 + \tfrac{1}{64}l_{|ab}^2$$

Variance of BIP means,

$$V_{1S3} = \tfrac{1}{4}d_a^2 + \tfrac{1}{4}d_b^2 + \tfrac{1}{16}h_a^2 + \tfrac{1}{16}h_b^2 + \tfrac{1}{16}i_{ab|}^2 + \tfrac{1}{64}j_{a|b}^2 + \tfrac{7}{64}j_{b|a}^2 + \tfrac{1}{256}l_{|ab}^2$$

Mean variance of BIP families,

$$V_{2S3} = \tfrac{1}{4}d_a^2 + \tfrac{1}{4}d_b^2 + \tfrac{3}{16}h_a^2 + \tfrac{3}{16}h_b^2 + \tfrac{3}{16}i_{ab|}^2 + \tfrac{7}{64}j_{a|b}^2 + \tfrac{7}{64}j_{b|a}^2 + \tfrac{15}{256}l_{|ab}^2$$

Covariance of F_2 and BIP means,

$$W_{1S23} = \tfrac{1}{4}d_a^2 + \tfrac{1}{4}d_b^2 + \tfrac{1}{16}i_{ab|}^2$$

BIP stands for biparental families of the third generation. This generation is referred to as S_3.

form of terms of the type $(d - \tfrac{1}{4}j)^2$ and $(h - \tfrac{1}{4}l)^2$. The size of these terms will obviously depend on the sign, and hence direction of the interactions. The partial pooling of interaction with main effect could serve either to enhance or to diminish the contribution made to the statistics according to the direction of interaction. Where several interacting genes are involved in the system some terms might, of course, tend to enhance, and others to diminish, the variance simultaneously.

The same is also true of the contributions to the summed variances of the backcrosses, though the compound terms are different, involving a different j with a given d, and i instead of l with h. The remainder of the interaction effects appear in compound terms involving i with l and the two j's together. There is a further complication in these backcross variances, for the size of each compound term varies with

the distribution of the genes between the parents. The backcross variances are indeed subject to so many sources of complication that they are likely to be relatively uninformative.

The F_3 and S_3 statistics should be informative in different ways. Since the interactions remain unconfounded in the S_3 statistics, they can be used to help directly in the separation of main and interactive effects. The covariance, W_{1S23} , is likely to be of special value as it includes only terms in d^2 and i^2 and so provides in a sense a direct measure of the fixable heritable variance since i is the fixable interaction. Statistics from later S generations are not likely to have this same advantage, as they will almost certainly contain terms in which parts of the interactions are confounded with main genic effects.

The F_3 statistics already show this confounding of the interactions, and they enable us to see something of its effects. The two variances, V_{1F3} and V_{2F3} , contain the same types of term as each other, though with different coefficients. The terms in W_{1F23} are the geometric means of the corresponding terms in V_{1F3} and V_{1F2} . Thus the term $\frac{1}{2}d_a^2$ in V_{1F2} is replaced by $\frac{1}{2}(d_a - \frac{1}{4}j_{a|b})^2$ in V_{1F3} and the corresponding term in W_{1F23} is $\{\frac{1}{2}d_a^2 . \frac{1}{2}(d_a - \frac{1}{4}j_{a|b})^2\}^{\frac{1}{2}} = \frac{1}{2}d_a(d_a - \frac{1}{4}j_{a|b})$. The corresponding term in V_{2F3} depends on $(d_a - \frac{1}{4}j_{a|b})^2$ just as in V_{1F3} , but of course, with its own characteristic coefficient. To put all this in another way, the definitions of D and H change from $\sum(d^2)$ and $\sum(h^2)$ respectively in F_2 to $\sum(d - \frac{1}{4}\sum j)^2$ and $\sum(h - \frac{1}{4}\sum l)^2$ in F_3 , so that the basic constitution of the terms, or components, of variations changes with generations but is constant over ranks within the generation. The summation sign is placed before j and l to indicate summation over all the appropriate digenic interactions which this gene shows with its fellow members of the polygenic system. That this is a general property can be seen from the general formulae for the variance of rank m in generation n of the selfing series and the covariance of rank m in generations n and n' which are

$$V_{mFn} = 2^{-m} \sum_a \left(d_a - \left(\tfrac{1}{2} - 2^{-n+1}\right) \sum_b j_{a|b}\right)^2$$

$$+ 2^{-2n+m+1} \sum_a \left(h_a - \left(\tfrac{1}{2} - 2^{-n+1}\right) \sum_b l_{|ab}\right)^2$$

$$+ 2^{-2m}(2^{m+1} - 3) \sum_{a<b} i_{ab|}^2$$

$$+ 2^{-2n-m+2}(2^{2m-1} - 1) \sum_{a,b} j_{a|b}^2$$

$$+ 2^{-4n+m+4}(3.2^{m-2} - 1) \sum_{a<b} l_{|ab}^2$$

and $\quad W_{mFnn'} = 2^{-m} \sum_a \left(d_a - (\tfrac{1}{2} - 2^{-n+1}) \sum_b j_{a|b} \right)$

$$\cdot \left(d_a - (\tfrac{1}{2} - 2^{-n'+1}) \sum_b j_{a|b} \right)$$

$$+ 2^{-n-n'+m+1} \sum_a \left(h_a - (\tfrac{1}{2} - 2^{-n+1}) \sum_b l_{|ab} \right)$$

$$\cdot \left(h_a - (\tfrac{1}{2} - 2^{-n'+1}) \sum_b l_{|ab} \right)$$

$$+ 2^{-2m}(2^{m+1} - 3) \sum_{a<b} i_{ab|}^2$$

$$+ 2^{-n-n'-m+2}(2^{2m-1} - 1) \sum_{a,b} j_{a|b}^2$$

$$+ 2^{-2n-2n'+m+4}(3.2^{m-2} - 1) \sum_{a<b} l_{|ab}^2$$

Now the definitions of D and H also change when the genes are linked, but they change with rank and not with generation (Table 5).

TABLE 5

Changes in the Main Components of Variation with Interaction and Linkage

	Statistic	Coefficient	Structure of Component		
			Simple	Interaction	Linkage
D	V_{1F2}	$\tfrac{1}{2}$	$\sum d_a^2$	$\sum d_a^2$	$\sum d_a^2 \pm 2 \sum d_a d_b (1 - 2p_{ab})$
	W_{1F23}	$\tfrac{1}{2}$	$\sum d_a^2$	$\sum d_a(d_a - \tfrac{1}{4} \sum j_{a|b})$	$\sum d_a^2 \pm 2 \sum d_a d_b (1 - 2p_{ab})$
	V_{1F3}	$\tfrac{1}{2}$	$\sum d_a^2$	$\sum (d_a - \tfrac{1}{4} \sum j_{a|b})^2$	$\sum d_a^2 \pm 2 \sum d_a d_b (1 - 2p_{ab})$
	V_{2F3}	$\tfrac{1}{4}$	$\sum d_a^2$	$\sum (d_a - \tfrac{1}{4} \sum j_{a|b})^2$	$\sum d_a^2 \pm 2 \sum d_a d_b (1 - 2p_{ab})^2$
H	V_{1F2}	$\tfrac{1}{4}$	$\sum h_a^2$	$\sum h_a^2$	$\sum h_a^2 + 2 \sum h_a h_b (1 - 2p_{ab})^2$
	W_{1F23}	$\tfrac{1}{8}$	$\sum h_a^2$	$\sum h_a(h_a - \tfrac{1}{4} \sum l_{|ab})$	$\sum h_a^2 + 2 \sum h_a h_b (1 - 2p_{ab})^2$
	V_{1F3}	$\tfrac{1}{16}$	$\sum h_a^2$	$\sum (h_a - \tfrac{1}{4} \sum l_{|ab})^2$	$\sum h_a^2 + 2 \sum h_a h_b (1 - 2p_{ab})^2$
	V_{2F3}	$\tfrac{1}{8}$	$\sum h_a^2$	$\sum (h_a - \tfrac{1}{4} \sum l_{|ab})^2$	$\sum h_a^2 + 2 \sum h_a h_b (1 - 2p_{ab})^2$ $\cdot (1 - 2p_{ab} + 2p_{ab}^2)$

p_{ab} is the frequency of recombination between genes $A–a$ and $B–b$. The $\pm$ in the linkage terms of D indicates addition for coupling and subtraction for repulsion.

We have thus a means of separating the effects of interaction and linkage. In the absence of interaction, D and H are homogeneous over V_{1F2}, V_{1F3} and W_{1F23}, but will change in V_{2F3} where linkage is acting. With interaction, D and H will be inhomogeneous over V_{1F2}, V_{1F3} and W_{1F23}, and they will vary no more between these three statistics as a

group on the one hand and V_{2F3} on the other, than they do within the group of three. Thus the tests of residual interaction and linkage used by Mather (1949a) and by Mather and Vines (1952) are sound—with one proviso. Only the j and l interactions cause inhomogeneity among the rank one variances and covariance of F_2 and F_3. The definitions of D and H are unaffected by i interactions. These i interactions will not therefore be detected by the test of residual interaction, and may serve to inflate V_{2F3} as compared with the first rank statistics, since the coefficient of i^2 is disproportionately large in this second rank variance. Inflation of V_{2F3} would mimic repulsion linkage. The i interactions may, therefore, be confused with repulsion linkage, though they would never mimic coupling linkage in their effects. This constitutes the only danger of confusion when conclusions are based on data from F_2 and F_3 ; and the inclusion of further types of family may well afford a means of removing even this possible confusion. The final resolution of this remaining problem must, however, await the fuller consideration which is now being given to the effects of interaction on statistics from families in series obtained by mating systems other than selfing.

One point may perhaps be reiterated in conclusion. Our consideration applies to all types of digenic interaction, for, as we have seen, all such interactions, whether we would recognise them as of the complementary, epistatic or any other kind, can be represented, combined and manipulated in terms of i, j and l. The contributions made to the various means, variances and covariances by a pair of genes showing any of the classical types of interaction can be simply obtained as special cases from the general expressions of Tables 3 and 4 by imposing the appropriate relations between d, h, i, j and l from Table 2. And, finally, the present method of representation and analysis can be extended to trigenic and higher interactions with, we believe, equal prospects of successful understanding and interpretation.

Summary

The knowledge that the genes mediating continuous variation are carried in the nucleus enables us to assume the genetical specification of the families in suitably designed experiments, except in respect of linkage relations which must generally be inferred from the variation observed. The specification of phenotypic effect of the genes is, however, seldom if ever capable of the same precise assumptions. The effects of the genes, and their interactions, must generally be inferred from the phenotypic variation observed.

The phenotypic effect of a gene can be described completely in terms

of the parameters d and h used by Fisher *et al.* (1932) and by Mather (1949a). Four more parameters are required for the complete description of a digenic interaction. These may be conveniently defined as the "interaction" comparisons (in the statistical sense) of the d and h "main effects" of the two genes. Thus i_{ab} is the interaction of d_a and d_b, $j_{a|b}$ that of d_a and h_b, $j_{b|a}$ that of h_a and d_b and $l_{|ab}$ that of h_a and h_b.

All digenic interactions, including the classical types such as complementary action, epistatic action and so on, can be defined in terms of relations between d, h, i, j and l. Different types of interaction (in the classical sense) can thus be expressed and combined in terms of i, j and l. This system of describing interactions is capable of extension to trigenic and higher orders of interaction.

The effects of digenic interaction on means, variances, covariances and scaling tests derivable from backcrosses, F_2, F_3 and third generation biparental progenies (S_3) of a cross between two true breeding lines are analysed, and shown to be usefully expressible in terms of i, j and l. The use of scaling tests and of the second degree statistics in detecting digenic interactions is considered, and it is shown how the effect of interaction may be separated from that of linkage in the second degree statistics obtainable from F_2 and F_3. The only confusion to be anticipated is of i type interaction with repulsion linkage. Other types of family should help to remove even this possible confusion.

REFERENCES

Bateman, A. J. and Mather, K. (1951) The progress of inbreeding in barley. *Heredity*, *5*: 321–48.
Fisher, R. A., Immer, F. R., and Tedin, O. (1932) The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics 17*: 107–24.
Mather, K. (1949a) *Biometrical Genetics*. Methuen, London.
—— (1949b) The genetical theory of continuous variation. *Proc. 8th Int. Cong. Genetics. Hereditas. Suppl. Vol.* 376–401.
—— and Vines, A. (1952) The inheritance of height and flowering time in a cross of *Nicotiana rustica*. *Quantitative Inheritance*, 49–80 ed. E. C. R. Reeve and C. H. Waddington. H.M.S.O. London.

QUANTITATIVE STUDIES IN DIPHTHERIA PROPHYLAXIS: AN ATTEMPT TO DERIVE A MATHEMATICAL CHARACTERIZATION OF THE ANTIGENICITY OF DIPHTHERIA PROPHYLACTIC*

L. B. HOLT

The Wright-Fleming Institute of Microbiology,
St. Mary's Hospital Medical School,
Paddington, London, W.2.

Examined quantitatively, the antibody responses of animals and children to inoculations of different forms of diphtheria prophylactic vary greatly. The dose-response curves from such materials, however, do show a similar pattern, and there is a great variation between different forms of prophylactic in the dosage required to induce some arbitrary level of response (Jerne & Maaløe, 1949).

As Jerne & Wood (1949) point out, an assay of a test preparation (T.P.) in terms of a standard preparation (S.P.) is strictly valid only if "the less potent preparation behaves as though it were a dilution of the other in a completely inert dilutent. The relative potency of the T.P. in terms of S.P., defined as the ratio of doses required to produce a given response, is then independent of the dose level of response at which it is measured. They continue ". . . this is the only definition of relative potency that would normally be regarded by the bio-assayist as satisfactory . . . . An instance of current interest in which this assumption does not hold is the assay of diphtheria and tetanus toxoids in commercial products containing aluminium hydroxide, using as S.P. a reference sample of highly purified toxoid . . . the dose-response curves of the two preparations have different upper asymptotes and cannot be described by the same form . . . . The assay is thus invalid."

Here then we have the problem; two preparations have a property in common, viz. the ability to cause the development of antitoxin when injected, but the one cannot be expressed quantitatively in terms of the other in the usual way.

The present communication is concerned with (a) an attempt to overcome this difficulty by finding antigenicity equations applicable to all types of diphtheria prophylactic (and probably other antigens)

---

by which they can be completely described in mathematical terms, and
(b) indicating some difficulties involved in the translation of results
obtained in the laboratory to the field.

When we give groups of children or animals an inoculation of an
antigen, we may measure the specific response in two quite different
ways, (a) by the percentages that attain or exceed some arbitrary level
of response, or (b) by determining the geometric mean titre of responses.
Whichever method is used, it is known that the distribution of titres
among a group of similar subjects identically treated is lognormally
distributed: *vide* Barr (1950) in respect of horses, Barr, Glenny &
Randall (1950) for children, and Holt (1951) for guinea pigs. In practice
we are more often interested in knowing the percentage that fails to
attain some measure of response than to know the percentage at each
level of response. From this it follows that the table of the cumulative
normal distribution is of more value to us than that of the ordinate
(Fisher & Yates, 1948).

Since the distribution of titres in a group is lognormal, it follows
(a) that comparisons should be made in terms of geometric means or
log geometric means and (b) that two groups may have the same
geometric mean but the standard deviations of logs of titres may
differ considerably; therefore a strict comparison cannot be made simply
from the geometric means.

The relationship between the geometric mean titre and the percentage
that attains or exceeds some arbitrary titre may be expressed as

$$\overline{U} = \log u + \sigma \,(\text{probit } y - 5) \qquad\qquad\qquad \textbf{(a)}$$

where

$\overline{U}$ = log geometric mean titre

u = some arbitrary titre

σ = standard deviation of logs of titres

y = the percentage of subjects possessing u units of
antitoxin, or more, per ml. of serum.

If we now examine the results obtained from *a series of graded
inocula* (similar subjects and the same material and testing technique,
etc.), a dose response curve may be drawn by plotting the percentage
possessing some arbitrary level of antitoxin response, or more, against
the dose administered, and this is characteristically sigmoid in shape.
When, however, the probit of that percentage is plotted against the
logarithm of the dose administered, a straight line, i.e. the probit
regression line, may be obtained (Hazen, 1914; Whipple, 1916; Finney,
1952). The experimental evidence for this statement (e.g. Carlinfanti,

1948; Holt & Bousfield, 1949) relates to the probit of the Schick conversion rate (S.C.R.) which does not correspond precisely to the percentage, y, of subjects attaining an arbitrary titre (see later section). Under certain plausible assumptions, however, a linear relationship between probit S.C.R. and the log dose implies a linear relationship between the probit of y and the log dose (see the discussion on Schick Conversion below).

The equation for the straight line may be written

$$\text{probit } y = b \log Z + C$$

where

$$b = \text{slope of the regression line of probit } y \text{ on log dose,}$$
$$Z = \text{the dose,}$$
$$\text{and } C = \text{a constant.}$$

If, therefore, a dose z is required to give a 50% response, then the probit for the percentage attaining, or exceeding, the titre u from a dose Z will be

$$b(\log Z - \log z) + 5$$

which may be rewritten as

$$\text{probit } y = b \log \left(\frac{Z}{z}\right) + 5 \qquad (b)$$

In brief, *equation (a) describes the distribution of responses at one dose, whereas equation (b) describes the whole dose-response curve.*

Now equation (b) may be combined with equation (a) with the advantage of having all the variables present in one expression. Let

$$\overline{U}(Z) = \text{log geometric mean titre from a dose } Z$$

and

$$\overline{U}(z) = \text{log geometric mean titre from a dose } z.$$

Then by substituting the right-hand component of (b) for the term "probit y" in (a), and putting $\log u = \overline{U}(z)$ since z gives a 50% response, we obtain

$$\overline{U}(Z) = \overline{U}(z) + B \log (Z/z), \qquad (c)$$

where $B = \sigma b$.

The slope, B, of the regression line log geometric mean of titres on log dose is, therefore, equal to σb.

It is essential for the validity of this transposition that σ remains constant for all doses used for a given type or sample of stimulus.

The whole mathematical model may be described by the following comprehensive equation, relating the probit of y (the percentage of antitoxin titres exceeding an arbitrary value u) to u and z (the log dose):

$$\text{Probit } y = 5 + b \log (Z/d) - (1/\sigma) \log (u/u_0), \qquad \text{(d)}$$

where b and σ are defined as above and d is the dose required to give a geometric mean titre equal to some value u_0 .

It will be seen that equation (d) provides for the complete characterization of an antigen in physiological terms. It is a *sine qua non* that the animal, age or weight of animal, and number of doses, route of inoculation and time interval(s) employed must always be specified when values are given to the variables. Three constants enter into equation (d). These may be taken to be σ, b and the dose, d, required to produce some arbitrary geometric mean titre, which might be usefully taken as 0.003 *u./ml.* for diphtheria antitoxin in children, as this corresponds approximately to a 50% Schick conversion rate (see below).

The constants σ and b appear in equation (c) only through their product, B. I have suspected that the value of B is unity, Prigge (1953) contends that this must always be so, and indeed by applying the conversion formula (*infra*) for geometric mean from S.C.R. to the field data given by Holt & Bousfield (1949), B is estimated to be 0.9. Even if B were unity or some other constant it would still be necessary to determine σ and b separately for a complete characterization of antigenicity.

DISCUSSION

For a full characterization of the antigenic properties of samples of diphtheria prophylactic, we need to know the values of σ and b as determined in *children*, and the dose required to produce some arbitrary measure of response. This latter, however, may be very different for the same preparation and subject without evident alteration in the values of σ and b; for instance, Holt & Bousfield (1949) comparing the Schick conversion rates, in children, from P.T.A.P. (Holt, 1947) administered (a) subcutaneously and (b) intramuscularly, found that the regression lines of probit S.C.R. on log dose were parallel but differed considerably in position.

The values of σ, b and B have been determined for responses to a single injection of P.T.A.P. in guinea-pigs (Table I). The values of σ have been calculated for both primary and secondary responses for

TABLE I.
Antigenicity Constants for P.T.A.P. in Guinea Pigs: Single Dose.

B (Slope of log geometric mean titre on log dose) . . . . . . 0.6   (Holt, 1950)
σ (Standard deviation of log titres) . . . . 0.512 (see Table II)
b (= B/σ) 1.15

many batches of P.T.A.P., and their mean values and standard deviation are shown in Table II. It is of interest to note that σ does not alter greatly from one to two doses; this is in marked contrast to the effect of A.P.T. in guinea-pigs (Barr & Llewellyn-Jones, 1951).

TABLE II.
Data on the Standard Deviation of Logs of Titres, σ, in Guinea Pigs, for P.T.A.P.

	Primary Responses	Max. Secondary Responses
No. of Groups. (12 per group)	31	30
Mean Value of σ	0.512	0.436
Range of Values of σ	0.232 − 0.874	0.189 − 0.790
Standard Deviation of Values of σ	0.168	0.173

In respect of information from children the data available are not entirely satisfactory. In the following section an empirical formula is derived relating the log of the geometric mean titre to the S.C.R. viz.

$$\log \text{G.M.} = \bar{3}.5 + 0.7 \,(\text{probit S.C.R.} - 5) \qquad\qquad (e)$$

If b' is the slope of the probit S.C.R./log dose regression line, we should estimate B as 0.7 b'. From the data of Holt & Bousfield (1949) on P.T.A.P. the estimate of B is 0.9. Barr, Glenny and Randall (1950) give data for *two* doses of A.P.T. in which σ is approximately 0.4; this value is in close agreement with that calculated from other sources (see next section) where a mean value of 0.42 is found.

Manufacturers of diphtheria prophylactics in almost all countries are obliged to test (or have tested) all material intended for human use. The tests are carried out in guinea pigs and certain minimal requirements of antigenicity have to be fulfilled in order that the material be admitted as sufficiently potent (e.g. British Therapeutic Substances Regulations, 1952, and the National Institutes of Health (U.S.A.) requirements, 1948).

The implication of these requirements is that tests on the guinea pig may, with reasonable safety, be used as a substitute for tests on children, and in a broad measure this is so (W.H.O. Technical Report No. 61). But recently results have come to light which cast some doubt on the reliability of the guinea pig in this kind of work. The position is made more serious by the proposed adoption of "Standard Antigens" (W.H.O. Technical Reports Nos. 36 and 61) which in itself is, of course, very desirable. As we have already seen (Jerne & Wood, 1949) one cannot express the antigenicity of aluminium hydroxide absorbed toxoid in terms of purified toxoid in simple solution, although it may be practicable to have "Standard Antigens" against which to standardise broadly similar types of material.

The discrepancies found between laboratory (guinea pig) data and field (child) data are as follows:

I. Using guinea pigs and comparable doses Barr & Llewellyn-Jones (1951) found that the value of σ following a single injection of A.P.T. was much greater than that following P.T.A.P., and, in addition, the geometric mean titre of antitoxin was about six times greater with P.T.A.P. than with A.P.T. Holt & Bousfield (1949) using year-old children found that one dose of A.P.T. gave an 87.5% S.C.R. and one dose of P.T.A.P. a 95-97% S.C.R. If, in the children, the geometric mean titre from P.T.A.P. were six times greater than that from the A.P.T. and in addition equation (e) held in both cases, then the S.C.R. from the A.P.T. would not have exceeded 80%.

II. The second discrepancy would seem to be more marked than the first.

When *H. pertussis* vaccine is added to purified toxoid in solution, and comparative antigenicity tests made in guinea pigs it is found that the whooping cough vaccine has considerably augmented the response to the toxoid component of the mixture, measured by their responses to one or two doses (Faragó & Pusztai, 1949; di San't Agnese, 1949; Ungar, 1952). Bousfield & Holt (1953) found that their vaccine increased the antitoxin responses in guinea pigs some 12-15 fold for a single inoculation. In children the same toxoid alone gave a 63% S.C.R. and the toxoid-vaccine mixture an 83.8% S.C.R. (Table III). From equation (e) this increase in S.C.R. indicates an increase of about threefold in geometric mean titre of antitoxin in the children. If the difference in log geometric means which was found in the guinea pig data had been directly transferable to children then the S.C.R. would have been about 96%.

All the above difficulties may be avoided and the assessment of the

antigenicity characteristics be accurately determined by specifying the three constants in equation (d). In practice this means that a part of the dose-response curve for the prophylactic under test must be measured *in children*; two dosages having a 5:1 ratio, with 30-50 children at each, would probably be adequate. From an examination of the

<div align="center">TABLE III.</div>

Comparative Antigenicity of Purified Diphtheria Toxoid Alone and Mixed with H. Pertussis Vaccine, in Guinea Pigs and in Children (Bousfield & Holt, 1953).

GUINEA PIG DATA

	Dosage	Geometric Mean Titres *U/ml.*	
		Single dose	Two doses
Exp. 1.			
(a)	1.4 Lf toxoid	0.012	0.242
(b)	1.4 Lf toxoid plus		
	400 M. H. pertussis	0.191	5.76
	Ratio (b)/(a)	16	24
Exp. 2.			
(a)	1.0 Lf toxoid	0.0036	0.058
(b)	1.0 Lf toxoid plus		
	285 M. H. pertussis	0.042	2.85
	Ratio (b)/(a)	12	49

CHILD DATA (S. C. R. Measured 4 weeks after a single dose)

(a)	30 Lf toxoid	gave 63%	S.C.R. (213 cases)
(b)	30 Lf toxoid plus 10,000 M.H. pertussis	gave 83.8% S.C.R. (213 cases)	

individual results obtained equation (d) could be completed. The bleeding of small children for assessment of antibody responses is becoming increasingly practiced to-day (e.g. Butler, Barr & Glenny, 1954).

Such work need only be done on the "Standards" proposed by the B.S. Committee (W.H.O.) and on new prophylactics as they are developed. The field-calibrated standards may then more reliably be employed in the laboratory for routine work.

A NOTE ON THE SCHICK NEGATIVE REACTION RATE.*

Many workers have observed that there is no one clear-cut serum antitoxin titre at which all subjects pass from a state of giving a Schick positive reaction to a negative one. Nevertheless all investigators agree that the higher the mean titre of a group the higher is the negative reaction rate in that group (Leach 1935; Parish & Wright 1938; Downie *et al* 1941; Greenberg & Roblin 1949).

In field trials where large groups of children are employed and the Schick test is used as the indicator of prophylactic efficiency, it would be of value to be able to translate percentage Schick conversion rate into geometric mean antitoxin titre.

Much of the published work in the relationship between serum antitoxin titre and the Schick test result is not valid as the reagent used (Test Toxin) was not standardised (League of Nations B.S. Report 1931; British T.S.A. Regulations, 1931) or the test and bleeding were not made simultaneously. The quantity of published data is still further restricted by the serum antitoxin titres being recorded as having a potency greater than or alternatively less than some value.

The method of using the available data (Table IV) was to calculate the geometric mean of the extremes of the titration brackets used by the authors, and plot the percent negative reactors in that group against that titre. The percent Schick negative reaction rate (S.N.R.R.) increased in a sigmoid curve to a 100% asymptote with increase of geometric mean.

When the probit of the percent negative reaction rate was plotted against log geometric mean, a straight line could reasonably be drawn through the points. The probit line was fitted by the standard method (Finney, 1952), and a slope of 1.435 ± 0.159 obtained. The test for heterogeneity gave $\chi^2 = 8.667$ for 5 d.f. $(P = 0.15)$. The geometric mean titre corresponding to a 50% S.N.R.R. was estimated to be 0.0032 $U/ml.$, whence, as a first approximation, the relationship between S.N.R.R. (or S.C.R.) and geometric mean titre may be expressed as

$$\text{log geometric mean titre} = \bar{3}.5 + \frac{1}{1.43} (\text{probit S.N.R.R.} - 5),$$

which is equivalent to equation (e) above.

The data provided by Downie *et al.* (1941) were found to be completely at variance with the remainder. These particular data were

---

*The expression "% S.C.R." is customarily used to describe the antigenic efficiency of a course of active immunization. The expression "Schick Negative Reaction Rate" as used here, is meant to focus attention on the physiology of the response to the Schick test as distinct from the immunizing reagent. Mathematically these expressions are interchangeable.

obtained at the height of the secondary response (10 days after a second inoculation) whereas the others (apart from data B, 2nd figures) were derived from subjects which had received their injection(s) at least two months before the tests were made.

Recently, Kurokawa *et al.* (1951) reported a marked discrepancy in the relationship between circulating antitoxin titre and the Schick test

TABLE IV.

Collected Data on the Relationship between Geometric Mean Titre of Serum Antitoxin and the Schick Test Results in Children.

Calculation of the Slope of the Regression Line of Probit S.N.R.R. on log G.M. and the value of χ^2 for Heterogeneity.

Origin*	Log G.M. Titre	Cases	No. — ve	% — ve	Probit % — ve
A.2	$\bar{2}.35$	6	6	100	∞
A.1		57	50	87.7	
C	$\bar{2}.24$	22	21	95.4	
A.1 + C.		79	71	89.9	6.28
B.	$\bar{2}.15$	9 + 20	7 + 15	75.9	5.70
D.	$\bar{2}.00$	27	—25	92.6	6.45
A.1		16	9	56.2	
B.		5 + 17	3 + 5	36.4	
C.	$\bar{3}.85$	11	11	100	
A.1 + B. + C.		49	28	57.1	5.18
D.	$\bar{3}.60$	25	24	96	6.75
A.2		17	13	76.5	
B.	$\bar{3}.50$	16 + 17	10 + 3	39.4	
A.2 + B.		50	26	52	5.05
D.	$\bar{3}.30$	22	19	86.4	6.1
A.1		26	10	38.5	
C.	$\bar{3}.20$	21	14	66.7	
A.1. + C.		47	24	51	5.02
D.	$\bar{3}.0$	12	7	58.3	5.21
A.2		48	11	22.9	
B.	$\bar{3}.0$	20 + 16	5 + 3	22.2	
A.2. + B.		84	19	22.6	4.25

Statistical Analysis

1) Slope $= 1.435 \pm 0.159$
   $\chi^2_{(5)} = 8.667, P = 0.15.$ } Data D excluded.

2) σ for group

   $A_1 = 0.44$ $B(2) = 0.43$

   $A_2 = 0.40$ $C = 0.46$

   $B(1) = 0.44$ $D = 0.35$

   Mean $\sigma = 0.42.$

result in guinea pigs in which (a) the serum antitoxin titre was rapidly
rising and (b) where it was uniform.   These observations may account
for the anomalous data of Downie *et al.*

The finding that there is a linear relationship between probit
S.N.R.R. and log geometric mean titre reveals the fact that there is
another variable operating with a distribution σ_0 .   The reciprocal of
the slope of the regression line of probit S.N.R.R. on log geometric mean
titre must, therefore, be the resultant of this other variable and σ in the
groups examined.   Presumably σ_0 is reasonably constant and represents
the variation in human skin capillary permeability.   The groups of
subjects examined all have very much the same value of σ (Table IV),
with a mean of 0.42.   Assuming that the skin capillary variable is also
lognormally distributed and is independent of the antitoxin titre, then
the reciprocal of the slope of the regression line of probit S.N.R.R. on
log geometric mean titre may be expressed as

$$0.7 = \sqrt{\sigma^2 + \sigma_0^2}$$

and since σ is, approximately, 0.42 the value of σ_0 is about 0.56.

SUMMARY

In view of the known gross differences in response from different
forms of diphtheria prophylactic an attempt has been made to character-
ise the antigenicity of any type in mathematical terms.

Use is made of the observation that the responses among a group
of similar subjects, identically treated, is lognormally distributed; as
well as the observation that when the probit of the Schick conversion
rate is plotted against log dose a straight line is obtained.

It is found that *three* variables are involved, namely d, the dose
required to effect some arbitrary measure of response; b, the rate of

increase, with respect to log dose, of the probit of the percentage of titres exceeding the arbitrary level; and σ, the standard deviation of logs of titres. In addition, it is found that the product of b and σ is equal to B, the slope of the line relating the log geometric mean to log dose.

The fact that σ may be different from different types of prophylactic signifies that neither the comparison of geometric means nor the determination of b can provide a strictly accurate characteristization of antigenicity; this has direct relevance to the use of standard or Reference Preparations for routine laboratory purposes.

It is suggested that the values of d, b and σ should be determined, *in children*, for each type of prophylactic, and in view of serious discrepancies between laboratory and field data that Standard Antigens should be calibrated in the field before adoption in the laboratory.

An examination of selected published data on the relationship between serum antitoxin titre and Schick test result is made. From these data a first approximation of the relationship between percent Schick negative and geometric mean titre of antitoxin has been derived, which may be expressed as

$$\log \text{G.M.} = \bar{3}.5 + 0.7 \text{ (probit percent negative} - 5).$$

The value of 0.7 for the reciprocal of the slope of the probit regression line, appears to be the resultant of two constants, σ the standard deviation of logs of titres and σ_0 the standard deviation of the distribution of skin capillary permeability in children, in that

$$0.7 = \sqrt{\sigma^2 + \sigma_0^2}$$

It is suspected that σ_0 is relatively constant and that σ varies with the type of diphtheria prophylactic used.

———————

REFERENCES

Barr, M. (1950) Brit. J. Exp. Path., *31*, 615.
Barr, M., Glenny, A. T., and Randall, K. J. (1950) Lancet *i* 6. Table I. Group A + B + C.
Barr, M. and Llewellyn-Jones, M. (1951) Brit. J. Exp. Path., *32*, 231.
Butler, N. R., Barr, M. and Glenny, A. T. (1954) B. M. J. *i* 476.
Bousfield, G. and Holt, L. B. (1950) B. M. J. *ii* 73.

Bousfield, G. and Holt, L. B. (1953) VI International Congress of Microbiology, Rome.

Carlinfanti, E. (1948) J. Immunol., *59*, 1. and Lancet *ii* 629.

Downie, A. W., Glenny, A. T., Parish, H. J., Wilson Smith & Wilson, G. S. (1941) B. M. J. *ii* 717.

Faragó, F. and Pusztai, S. (1949) Brit. J. Exp. Path., *30*, 572.

Finney, J. D. (1952) "Probit Analysis" 2nd Edition. Camb. Univ. Press.

Fisher, R. A. and Yates, F. (1948) "Statistical Tables" Table 1. Oliver & Boyd.

Greenberg, L. and Roblin, M. (1949) Canad. J. Public Hlth. March. 112.

Hazen, A. (1914) Trans. Amer. Soc. Civ. Engrs. *77*, 1539.

Holt, L. B. (1947) Lancet *i* 286.

Holt, L. B. (1950) "Developments in Diphtheria Prophylaxis" p. 154. Heinemann.

Holt, L. B. (1951) Brit. J. Exp. Path., *32*, 157.

Holt, L. B. and Bousfield, G. (1949) B. M. J. *i* 695.

Jerne, N. K. and Maaløe, O. (1949) Bull World Hlth. Org. *2*, 49.

Jerne, N. K. and Wood, E. C. (1949) *Biometrics*, *5*, 273.

Kurokawa, M., Murta, R., Nakano, T., Yamada, T. and Kubata, K. (1951) Jap. Med. J., *4*, 369. Table 2.

Leech, C. N. and Poch, G. (1935) J. Immunol., *29*, 367.

Parish, H. J. and Wright, J. (1938) Lancet *i*, 882.

Prigge, R. (1953) Bull. Wld. Hlth. Org. *9*, 843.

di San't Agnese, P. A. (1949) Pediatrics, *3*, 20.

Ungar, J. (1952) Proc. Roy. Soc. Med., *10*, 674.

Valquist, B. and Hogstedt, C. (1949) J. Immunol., *62*, 277.

Whipple, G. C. (1916) J. Franklin Inst., *182*, 37 and 205.

PREDICTION EQUATIONS IN QUANTITATIVE GENETICS

ALAN ROBERTSON

Institute of Animal Genetics, Edinburgh.

Member of the Scientific Staff of the
Agricultural Research Council, Great Britain.

One of the fundamental concepts in the application of statistical methods to the analysis of the inheritance of characters showing continuous variation is the additive genetic variance σ_g^2, the variance in any character in a population that is due to the average effects of genes. If this is expressed as a fraction of the total variance, σ_p^2, we get the related parameter, h^2, the heritability (in the narrowest sense) of the character. It can be shown without great labour that the heritability is also equal to the regression coefficient of breeding value on performance or phenotype. This short paper presents an alternative derivation from the point of view of the combination of information from different sources, an approach which may be useful in teaching. Several other important prediction equations in quantitative genetics can be fitted into the same pattern.

If we have a measurement P of an individual in a population in which the character measured has a mean $\overline{P}$, we may consider ourselves as having two independent pieces of information on the animal's breeding value. They are: (i) that the animal is a member of a population whose mean breeding value is $\overline{P}$ with variance σ_g^2; (ii) that the animal's own performance is P and that this will have variance $\sigma_p^2 - \sigma_g^2$ about the true breeding value.

If we knew only (i), we should take $\overline{P}$ as the best estimate of the individual's breeding and it would have error variance σ_g^2. If we knew only (ii), we should take P as the best estimate with error variance $\sigma_p^2 - \sigma_g^2$.

In combination, the correct weight to give the two estimates is the reciprocal of their respective variances. We then have

$$G = \frac{\dfrac{\overline{P}}{\sigma_g^2} + \dfrac{P}{\sigma_p^2 - \sigma_g^2}}{\dfrac{1}{\sigma_g^2} + \dfrac{1}{\sigma_p^2 - \sigma_g^2}}$$

$$= \frac{\overline{P}(\sigma_p^2 - \sigma_g^2) + P\sigma_g^2}{\sigma_p^2}$$

$$= \overline{P} + (P - \overline{P})\frac{\sigma_g^2}{\sigma_p^2}$$

$$= \overline{P} + h^2(P - \overline{P})$$

which is the usual regression formula. This derivation shows clearly the premises on which the formula is based.

We may now extend the scope of this presentation so that we can easily deal with other prediction formulae. The general situation is that we are interested in a primary variable whose probable value we wish to predict (in the previous case the breeding value), which is obscured by some secondary variation, which may itself be partly genetic in origin. The regression coefficient in the prediction formula is just the fraction which the real variance of the primary variable makes up of the total.

As a first example, we may wish to evaluate the breeding value of a series of males by a progeny test in which there is no environmental variance common to members of a progeny group, a condition which is perhaps not often fulfilled. The primary variance between groups due to sires is $\sigma_g^2/4$ while the secondary obscuring variation is due to sampling within groups and is equal to $[\sigma_p^2 - (\sigma_g^2/4)]/n$ where n is the number of offspring in the group. The regression coefficient in the prediction of the true value of progeny of this sire from the observed mean value of his n offspring is

$$\frac{\dfrac{\sigma_g^2}{4}}{\dfrac{\sigma_g^2}{4} + \dfrac{\sigma_p^2 - \dfrac{\sigma_g^2}{4}}{n}}$$

which on manipulation becomes the accustomed formula

$$\frac{\tfrac{1}{4}nh^2}{1 + \tfrac{1}{4}(n - 1)h^2}.$$

The information put into the prediction is (i) that the sire belongs to a given breed in which the genetic variance between progeny groups is

$\sigma_o^2/4$ (ii) that the observed average of his progeny has sampling variance $[\sigma_p^2 - (\sigma_o^2/4)]/n$.

The same formula would apply if we wished to predict the breeding value of another member of the progeny group whose own performance had not been included in the group average i.e. in family selection. We may cast the problem of family selection more generally as follows. Suppose we are dealing with a population made up of families of average relationship r in which the observed phenotypic correlation between relatives is t (see Lush, 1947). In other words, the genetic and pheno-typic components of variance within and between groups are

	Between groups	Within groups
Phenotypic	$t\sigma_p^2$	$(1-t)\sigma_p^2$
Genetic	$r\sigma_g^2$	$(1-r)\sigma_g^2$

Ignoring for the moment selection within families, we can take first the situation where the animal chosen is not itself measured, as for example when a cockerel is chosen on the egg production of his sisters, the members measured being considered only as representatives of the family. The regression is then given by

$$\frac{r\sigma_g^2}{t\sigma_p^2 + \dfrac{(1-t)\sigma_p^2}{n}} = \frac{nrh^2}{1 + (n-1)t}$$

If there is no environmental similarity between family members, we can write $t = rh^2$ and the formula becomes similar to that discussed above for progeny testing. If, on the other hand, we are choosing an individual whose measurement is included in the family average, we are interested in these actual groups of n relatives and the sampling contribution of the genetic variance must be included in the primary variance. Then we have for the regression coefficient

$$\frac{r\sigma_g^2 + \dfrac{(1-r)\sigma_g^2}{n}}{t\sigma_p^2 + \dfrac{(1-t)\sigma_p^2}{n}} = h^2 \frac{1 + (n-1)r}{1 + (n-1)t}$$

If we wish to take into account also the animal's own phenotype, it is simpler to use the two quantities $P - F$, the deviation of the individual from the family mean, and $F - \bar{P}$, the deviation of the family mean from the population mean. These have the advantage

that they are statistically independent and knowledge of one tells nothing about the other. We can then simply add together the predictions from the two variables. The regression coefficient of breeding value on family mean we have just obtained. As we are dealing with deviations from the observed mean, the effective genetic variance within families is $(1 - r) \sigma_o^2 (n - 1)/n$ and the phenotypic $(1 - t)\sigma_p^2 (n - 1)/n$. The regression coefficient for $P - F$ is then $h^2(1 - r)/1 - t$. The full equation reads

$$G = \overline{P} + h^2 \frac{1 + (n - 1)r}{1 + (n - 1)t} (F - \overline{P}) + h^2 \frac{1 - r}{1 - t} (P - F)$$

as given by Lush. This derivation of the basic equation of family selection is more congenial to the author than that by path coefficients and is presented as an alternative which may be of value to those learning the subject.

Summary

The basic prediction equation of quantitative genetics (that of breeding value on performance) is derived from the point of view of the combination of information from different sources. The principle is extended to several other prediction equations in family selection and progeny testing.

REFERENCES

Lush, J. L. (1947). Family merit and individual merit as bases for selection. *Amer. Nat. 81*, 241–261 and 362–379.

DETERMINING THE FRUIT COUNT ON A TREE BY RANDOMIZED BRANCH SAMPLING*

Raymond J. Jessen

Iowa State College
Ames, Iowa

In crop estimation work and in some areas of biological and pomological research, the problem of determining the total number of fruits on a tree sometimes arises. If an accurate count of all fruits is attempted, this may be quite an onerous and time consuming job,—especially, if the fruits are to be left on the tree undamaged. If the fruits are picked before counting, in order to improve the accuracy of the results, the removal of the fruits may seriously interfere with other aspects of the investigation. A method of obtaining reasonably precise estimates of the total fruits by sampling, so that little time is required, may be of some interest. The purpose of this paper is to describe some possible schemes and compare some aspects of their efficiencies.

The object of sampling is to select some portion of a relatively large total which will represent that total reasonably well. In the present case, the object is to select a few of the many smaller branches of a tree in such a manner that counting the fruits on these sample branches will enable us to obtain a reasonably accurate estimate of the total fruits on the tree. At present, we shall consider only those schemes which select the sample branches by a randomizing procedure.

Suppose the branching system of a tree is represented as in the following diagram:

The trunk, branch number "0", splits into two branches at fork I. Branch 1 of this fork splits into 3 branches at fork II, etc. Suppose all the fruits of the tree are borne on the peripheral branches, the number being indicated by the encircled figures. Thus branch 1 of fork III has 12 fruits, branch 1 of fork VI has none, etc.; the tree has 64 fruits borne on its 8 "fruiting" branches.

Suppose we wish to determine the fruit count of this tree by confining our counts to two fruiting branches selected at random. This could be done by numbering each of the 8 fruiting branches from 1 to
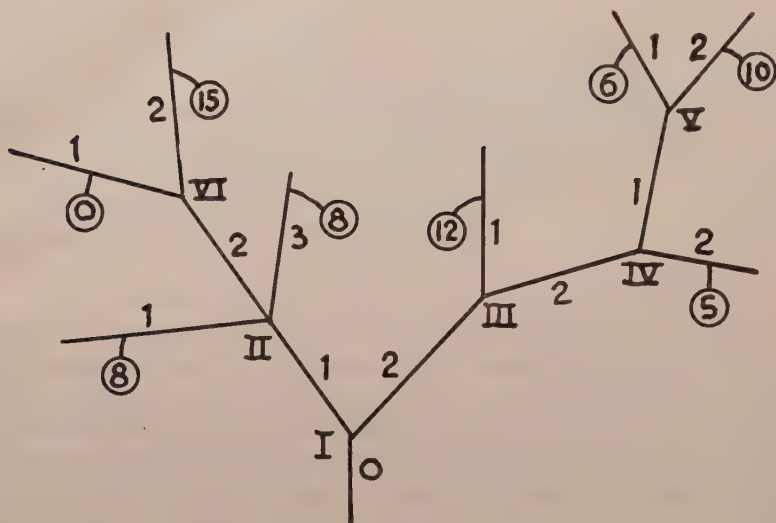
8, choosing two random digits from 1 to 8, say 8 and 4, and taking those designated branches (Nos. 8 and 4) for the sample. If counts are made on those two branches, we can obtain an average fruit count per branch, which when multiplied by the total number of fruiting branches, 8, provides an estimate of the total in the tree. If, in our case, serial number "8" refers to branch 2 of fork VI and serial number "4" is branch 2 of fork IV, we obtain the counts 15 and 5, an average of 10.0, or an estimated total for the tree of 8 × 10.0 = 80.

The above scheme is simple and, if counts are accurately made, will provide unbiased estimates of the total count. It may, however, be quite laborious to identify and number all the fruiting branches on a tree, such as this scheme requires, not only to provide a means for randomizing the selection of the branches but also to provide a means to estimate the total for the sample.

In order to avoid the problem of complete branch identification and numbering and still obtain unbiased estimates of the total fruit count, the following scheme is proposed. Let us take a position at fork I and decide by a random draw of a 1 or 2 whether to follow branch 1 or branch 2. Suppose 2 is drawn. We proceed up branch 2 to fork III and since there are two possible branches we draw another 1 or 2 at random, say 2 is drawn. Proceeding up to fork IV, suppose we draw another 2 at random which puts us at branch 2, our sample branch which must be counted. Obtaining the count of 5 fruits, we must now estimate the total on the tree, which is done as follows:

$$\text{Estimated total} = \frac{5}{1/2 \times 1/2 \times 1/2} = \frac{5}{1/8} = 40$$

The denominator of this estimator may be regarded as an estimate of the fraction of all fruiting branches that this particular sample branch represents. If two sample branches are desired, the above procedure can be repeated with new random draws. (If the same branch is selected, just repeat the second series of draws.) The estimate from the second branch is obtained in a manner identical to the first, and the best pooled estimate is simply the average of these two. For example, suppose on the second series we obtain branch 1 of fork V. The estimated total is given by

$$\frac{6}{1/2 \times 1/2 \times 1/2 \times 1/2} = 96$$

and the pooled estimate of the two sample branches is therefore

$$(1/2)(40 + 96) = 68$$

Although, in this example, the two-branch estimate, 68, is quite close to the true count, 64, it is more or less fortuitous. If all possible one-branch estimates are examined we obtain the following:

Fork and Branch No.	Branch Count	Estimate	Fork and Branch No.	Branch Count	Estimate
II-1	8	48	V-1	6	96
II-3	8	48	V-2	10	160
III-1	12	48	VI-1	0	0
IV-2	5	40	VI-2	15	180

It can be seen that our single branch estimates vary widely (from 0 to 180) depending on the particular branch selected. This undesirable characteristic of this method of sampling might be reduced somewhat by taking branch size into account in the scheme for selecting branches. Another alternative is to count a group of fruiting branches. These and other possible procedures for increasing the precision of the estimates for a given fraction of fruits counted will be dealt with later in this paper.

It may be of interest to test the unbiasedness of this method of estimating total fruits from branch samples. By unbiasedness is generally meant that the average of the estimates over all possible samples will be identical to the number being estimated. In the table

above we have the 8 possible estimates from single branch samples. Since the probability of obtaining a particular branch in a sample is not the same for all branches, we cannot take a simple mean of these 8 values as the average estimate of this method of sampling. A weighted average is required. The data for each of the 8 branches, the estimates obtained from each and the probability of obtaining each are:

Branch:	II-1	II-3	III-1	IV-2	V-1	V-2	VI-1	VI-2
Estimate:	48	48	48	40	96	160	0	180
Probability:	8/48	8/48	12/48	6/48	3/48	3/48	4/48	4/48

Weighting each estimate by its probability of occurrence we obtain as the weighted average, 64, which is identical to the true total being estimated. This scheme of sampling and estimating is therefore regarded as unbiased.

In order to provide an elementary test of the practicality of this scheme and to investigate the effects of certain modifications which seemed of interest, complete data were obtained from an orange tree. The tree, a pineapple orange approximately 25 years old, was situated in a Florida Citrus Experiment Station's experimental grove at Lake Alfred, Florida.* The counts were made September 23 and 24, 1953. The number of fruits borne on each of the branches was counted and recorded. The circumference of each branch (except the smaller ones) was measured near the fork of origin and also recorded. A total of 1379 fruits was counted. The results of the branch counts and measurements are shown in Figure 1.

The data collected in this manner provided the means for testing the efficiency of a number of alternative ways of selecting the sample branches. For example, what is the best basis for determining the probabilities with which to draw a branch at a given fork: (I) equal for each branch, (II) proportional to the number of branches into which each of these branches divide or (III) proportional to the cross-sectional area of each branch? To examine this question the 5 main branches were considered. It can be seen from Figure 1 that the trunk divides into two branches, say I and II, with circumference measurements 21 5/8″ and 23 1/2″ respectively, and I divides into two further branches, A with a circumference of 15″ and B with 11 7/8″; branch II on the other hand divides into 3 branches, A, B and C, with circumferences of 10 3/4″, 17 3/8″ and 14 1/2″ respectively. The three bases for determining probabilities are described as follows:

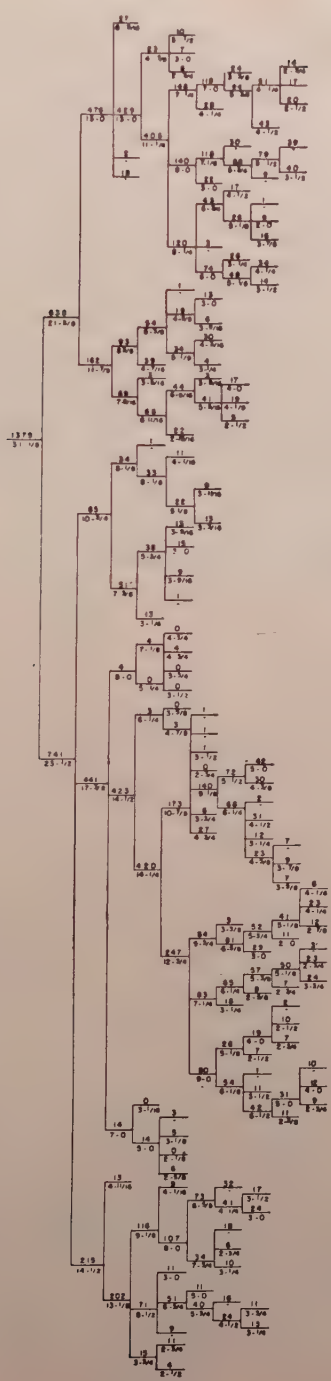*John W. Sites, granted permission for making the count.

FIGURE 1

PE   (probabilities equal). Since in this case there are 5 branches, each will have a probability of 1/5.

PPN   (probabilities proportional to "number"). With this scheme the probability of obtaining each main branch is made equal at each fork. In this case, since there are 2 branches at the first fork, the probability of each is 1/2. Apply the same principle at each of the subsequent forks, we obtain as the overall probability of getting branch IA, $1/2 \times 1/2 = 1/4$; for IB, $1/2 \times 1/2 = 1/4$; for IIA, $1/2 \times 1/3 = 1/6$; etc.

PPA   (probabilities proportional to "area"). As a measure of the cross-sectional area of a branch, the square of its circumference will be used. This scheme provides at any fork that large branches will have a greater chance of selection than a small branch. The following calculations are required for the first fork:

			Totals
Branch:	I	II	
Circumference:	21 5/8″	23 1/2″	
Circumference squared:	467.64	552.25	1019.89
Fraction of total, or prob.:	.46	.54	1.00

When similar calculations are carried out for the forks at the ends of these branches, we obtain as the final PPA, probabilities for the 5 branches:

Branches at first fork:	I		II		
Branch probabilities at first fork:	.46		.54		
Branches at second fork:	A	B	A	B	C
Branch probabilities at second fork:	.61	.39	.18	.48	.34
Overall probabilities:	.28	.18	.10	.26	.18

The extension of the foregoing procedures to the determination of selection probabilities for each scheme to any and all branches on the tree can be seen. For evaluating the effectiveness of the three procedures for selecting sample branches, we shall compare the variabilities of the estimates of total fruits in the tree obtained from each since the estimates will be made by the formula

$$\hat{X} = \frac{x}{P}$$

where $\hat{X}$ is the estimated number of fruits on the tree,

x is the actual number of fruits on a sample branch,

P is the probability of selecting the sample branch.

As a measure of the precision of the estimates, we may use the standard error of x or its square, the variance, which for samples of size one, is given by the formula:

$$V(\hat{X}) = \sum_{i=1}^{N} P_i(\hat{X}_i - X)^2$$

where $\hat{X}_i$ is the estimate obtained from one of the N different sample branches,

X is the true number of fruits—the quantity being estimated,

P_i is the probability of selecting the branch from which a particular estimate is made.

The variances corresponding to each of the three bases for selecting branches are shown here for the case where only the 5 main branches of the tree are regarded as sample branches. The relevant data are given in Table 1.

TABLE 1.

Data and comparisons of reliability of the three methods of sampling the 5 main branches of a tree.

	IA	IB	IIA	IIB	IIC	Totals
Branch Designation	IA	IB	IIA	IIB	IIC	
Branch Serial No., i	1	2	3	4	5	5
No. of fruits, x_i	476	162	85	441	215	1379
Prob. of selection, P_i ; PE	1/5	1/5	1/5	1/5	1/5	1.000
Prob. of selection, P_i ; PPN	1/4	1/4	1/6	1/6	1/6	1.000
Prob. of selection, P_i ; PPA	.28	.18	.10	.26	.18	1.000
Estimates, $\hat{X}_i$; PE	2380	810	425	2205	1075	
Estimates, $\hat{X}_i$; PPN	1904	648	510	2646	1290	
Estimates, $\hat{X}_i$; PPA	1696	903	874	1701	1171	
Variance, $V(\hat{X}_i)$; PE	—	—	—	—	—	602,115
Variance, $V(\hat{X}_i)$; PPN	—	—	—	—	—	597,224
Variance, $V(\hat{X}_i)$; PPA	—	—	—	—	—	128,545

In this case, the *PPA* method gave the greatest reliability, a variance of 128,545 as compared with 597,224 for *PPN* and 602,115 for *PE*, equal probability. Thus, it can be said that the efficiency of *PPA* relative to *PE* is 468% (= 602,115/128,545 × 100), a very clear superiority indeed.

The use of large branches such as these does not appear to be as generally practical for sampling as a smaller branch. By means of the same procedure given above, a comparison can be made of the respective efficiencies of branches of different sizes. It will be convenient to refer to "size" of branches by the average number of fruits on them. Thus, branches of two sizes, averaging about 17 and 25 fruits, will be compared for sampling efficiency, with branches averaging 276 fruits (the 5 main branches). The basic figures required for this comparison are the variances of estimates per branch made by the several methods. These figures are shown in Table 2.

TABLE 2.

Variances of estimates of total fruits on the tree for each of three sizes of sample branches and three selection schemes.

Branch description and total on tree	Size of branch (Average number of fruits per branch)	Variance of $\hat{X}$, estimated total fruits, per branch, by selection scheme:		
		PE (Probability equal)	PPN (Prob. prop. to No.)	PPA (Prob. prop. to area)
5 main branches	275.8	602,114	596,957	128,530
55 smaller branches	25.1	1,404,299	19,236,106	1,710,941
80 smallest branches	17.3	1,932,119	19,648,726	1,818,344

The variances in Table 2 were computed by the simple formula:

$$\sigma^2 = \sum_{i=1}^{N} \frac{x_i^2}{P_i} - X^2$$

where the tree has N possible sample branches, on some branch i the number of fruits is x_i, P_i is the probability of selecting the ith branch and X is the total number of fruits on the tree, the quantity being estimated. In order to compare the efficiency of the several methods, it is necessary to put them on a comparable basis. For example the variance of $\hat{X}$ when a 25.1 fruit branch is used is 1.4×10^6, and for a 17.3 fruit branch, it is 1.9×10^6, suggesting that the larger branch is more efficient for sampling. However, on the average, we must count 25.1/17.3 or 1.45 times more fruit with the larger branch. A comparison of the efficiencies of the two branch sizes can be made if the variances in Table 2 are put on a per fruit basis. This is equivalent to a comparison on the variances of the two schemes when the total number of fruits counted with each scheme is the same. The variances in Table 2

are put on a per fruit basis by multiplying each variance by the average number of fruit in the branch, K. Thus:

$$\sigma^2 \text{ (per fruit)} = K\sigma^2 \text{ (per branch)}.$$

In Table 3 are shown the variances on a per fruit basis and the corresponding relative efficiencies of the several schemes with the 17.3 fruit branch with equal probabilities of selection taken as a base.

TABLE 3.

Variances per fruit of estimates of total fruits on the tree for the several methods of sampling and the relative efficiencies of each.

Size of branch (Average numbers of fruits per branch)	Variance of $\hat{X}$, per fruit by selection scheme (in millions)			Relative efficiency of method; small branch both equal probability taken as 100		
	PE	PPN	PPA	PE	PPN	PPA
275.1	166.1	164.6	35.4	20.1	20.2	94.0
25.1	35.2	482.2	42.9	94.6	6.9	77.7
17.3	33.3	338.7	31.3	100.0	9.8	106.3

Under these conditions, the most efficient scheme is the small branch (17.3 fruits) selected with probability proportional to cross-sectional areas of the branches at each forking. A close second is the same small branch selected with equal probability. The least efficient is "middle" sized branch (25.1 fruits) selected with probabilities proportional to the numbers of branches at each forking. The expected loss in efficiency, as larger and larger branches are taken, shows up only when equal probability of selection is used. When other selection schemes are used, there is no clear trend. In general, the *PPN* scheme of selection is very poor, although it is probably the simplest and quickest to carry out when small samples are taken. Of the three probability schemes, the one using equal probability in selecting branches seems most difficult and time consuming to carry out in practice. Unless something better could be devised, it appears that each sample branch must be identified and given a number from 1 to N, so that branches can be selected purely at random. Operationally, the *PPA* is quite simple to carry out and gives no loss in efficiency over the simple random scheme.

The rather high efficiency of the large branches in the PPA scheme, should be regarded as spurious or, at least, with some skepticism. It must be remembered that it is based on only 5 observations, whereas the others are based on 55 and 80 observations, and all are based on one tree for one season!

With the PPA scheme of selection, it appears that stratification by main branch would not be very effective in increasing precision—particularly in view of the relatively high efficiency of the main branches as sample branches. Consequently, the indicated efficiency of the 5 main branches as strata with PPA selection of the 17.3 fruit branch gives an increase of only 8% over unstratified. This is probably smaller than that which could be expected from trees in general.

In the foregoing discussion, the assumption has been made that all fruits are borne on the "end" branches. In the case of oranges, a number of fruits are borne on small branches directly connected with relatively large branches. With the PPA scheme of selection, this "forking" can be dealt with as any other forking of branches. In this case, a relatively small probability is given to the selection of the small fruiting branch. However, this branch is usually so small in diameter that it is difficult to measure its relative size accurately. In this case, it may be advisable to count the fruits on this branch and then proceed up the tree with the sampling procedure. To obtain unbiased estimates we compute the estimate in two parts. For example, if we have made the following observations, where between the 2nd and 3rd forking 10 fruits were found and counted but sampling continued through the 5th forking where the sample branch yields 20 fruits:

Forking numbers:	1	2	3	4	5
Probability of branch, given the fork:	1/2	1/3	1/5	1/3	1/3
Number of fruits counted:			10		20

$$\hat{X} \text{ is given by:} \frac{[10]}{(1/2)(1/3)} + \frac{[20]}{(1/2)(1/3)(1/5)(1/3)(1/3)}$$

$$\text{or} \qquad 60 + 5400 = 5460.$$

In the foregoing work, the "intermediate" fruits (those which were counted along the sampling path as the 10 in the example) were combined with the sample branch fruits in the following manner:

$$\hat{X}_i = 5460$$

$$p_i = 1/2 \times 1/3 \times 1/5 \times 1/3 \times 1/3 = \frac{1}{270}$$

$$y_i = p_i\hat{X}_i = \frac{1}{270} \times 5460 = 20.222$$

fruits as compared with the corresponding "x_i" value of 20. The y_i, therefore, are the actual fruits to which an imputed value of the "intermediate" fruits is added. The 10 intermediate fruits can now be regarded as allocated to the sample branches and, in this case, the sample branch was allocated 0.222 fruits. In the analysis, the results of which are given in Tables 2 and 3, we actually used the y_i's instead of X_i's in order to keep constant the total number of fruits dealt with.

Summary and Conclusions

(1) A complete count of all fruits on an orange tree was made and the number found on each branch was recorded. Each branch having a circumference value of $1''$ or more near the forking was measured. A total of 1379 fruits was counted.

(2) Three methods of selecting branches as samples for estimating the total number of fruits were tested. Three different branch sizes were tested for efficiency.

(3) A method of selecting branches, wherein each branch at a forking is given a probability of selection proportional to its cross-sectional area, was found to be quite efficient. In fact, this scheme gave efficiency comparable to that in which each fruiting branch is selected with equal probability. The equal probability scheme is not practicable since it requires some identification of all fruiting branches before sampling can be carried out. The unequal probability scheme described herein does not require this information for unbiased estimates.

A FURTHER NOTE ON MISSING DATA

Horace W. Norton

Agricultural Experiment Station
University of Illinois

Nelder (1954) pointed out that an estimate of a missing datum is not merely a convenient value for facilitating analysis of variance, but is really an estimate of what would have been observed if the model on which that estimate is based is true. An error in his formula for the variance of the estimated missing value in a randomized block design should be corrected. The correct formula is

$$\text{var}(y) = \frac{r + t - 1}{(r - 1)(t - 1)} \, \sigma^2,$$

whereas Nelder has $(rt - 1)$ in the numerator.

It should prove helpful to some to point out that inspection suffices to show that Nelder's formula is incorrect. Remembering the *mathematical* model, it is obvious that the general mean, the constant for the affected block and that for the affected treatment can all be estimated with any desired accuracy, simply by increasing the numbers of blocks and of treatments. Hence so can their sum, which is the estimate of the missing value. Nelder's formula is not conformable with this observation, having a lower limit of σ^2 as r and t become large. On the other hand, his formula for the $r \times r$ Latin square is correct, and is of the order of $3\sigma^2/r$ as the square becomes large.

In referring to Query 96, which raised a question about "impossible" estimated values, another error has occurred in Nelder's paper. The missing value, estimated to be -6.64, has a sampling error of 8.23 on 32 degrees of freedom. The 95% confidence interval is therefore -6.64 ± 16.76 (rather than Nelder's value of 8.10), thus giving no appreciable indication whether the estimated value is based on an erroneous model.

There is some interest in the fact that not only missing values may have "impossible" estimated values. In the example of Query 96 the model leads to estimates of -3.23 and -1.48 for bait A for replications 4 and 11, respectively, but these are small compared with the sampling error of 8.23.

While tests of "possibility" of estimated values may occasionally prove useful, it is probably always better to test for additivity, as discussed for this example by Tukey (1954).

REFERENCES

Nelder, J. A., (1954) "A note on missing-plot values," *Biometrics*, *9:* 400–401.
Tukey, J. W., (1954) "Comment on Queries Nos. 96 and 103," *Biometrics*, *9:* 412–413.

QUERIES

George W Snedecor, *Editor*

QUERY: 113 In *Biometrics* 5, page 232 (1949), Tukey gave a test for additivity in a 2-way table. He indicated that the theory could be applied to other designs. We often make observations on animals in several periods of time using randomly selected Latin squares to allocate the animals to the periods. As an example, we counted responses of 5 animals, each subjected to 5 conditions, during 5 periods of one week each. The numbers of responses are shown in the table. How can we test additivity in this Latin square? (Note: The data are given in the first lines in Table I below. Ed.)

ANSWER: Let x denote the array of original observations. As in a simple 2-way table, the rows, columns and now treatments are bordered with means and deviations. The k-array contains constants due to fitting the additive model. For example, $k_{11} = 391.36 + 4.64 - 105.36 - 72.96 = 217.68$. Deviations $x - k$ are forced to add to zero in rows, columns and treatments.

Now form the y-array of Table II. The easiest one to use here is $y_{ij} = (k_{ij} - \bar{k} \cdots)^2$. As an example, $y_{11} = (217.68 - 391.36)^2 = 301,647$. It is convenient to divide each entry by 1000 then round. Except for the rounding, this will have no effect on the results. Analysis of variance of y gives $S = 533,996$, the interaction sum of squares.

Let $P = \sum y_{ij} (x_{ij} - k_{ij}) = (302)(-23.68) + \cdots + (2)(-12.88) = 9,232$. Then

$$\text{Sum of Squares for Non-additivity} = \frac{P^2}{S} = \frac{(9,232)^2}{533,996} = 160$$

We now have this analysis of variance:

Interaction, SS (Table I)	12	44,391	
Non-additivity	1	160	160
	—	——	
For Testing	11	44,231	4,021

Clearly, $F = 160/4,021$ is non-significant. There is no evidence against the hypothesis of additivity.

The example just given is the application to a Latin square of a general procedure, which can be applied to test nonadditivity in very general situations. In general, let x be the observations, k the result of fitting, and $x - k$ the residuals. Form $y = c (k - c_1)^2$, where c and c_1 are convenient constants (in the example $c = 0.001$ and c_1 was taken

x: Array of Original Observations, Conditions A, B, C, D, E.
k: Fitted Linear Model.
x − k: Deviations.

Animal		Period 1	Period 2	Period 3	Period 4	Period 5	Mean	Deviation
1	x	B 194	D 369	C 344	A 380	E 693		4.64
	k	217.68	404.08	396.48	342.48	619.28	396.00	
	$x-k$	−23.68	−35.08	−52.48	37.52	73.72		
2	x	D 202	B 142	A 200	E 356	C 473		−116.76
	k	204.08	174.88	192.08	383.68	418.28	274.60	
	$x-k$	−2.08	−32.88	7.92	−27.68	54.72		
3	x	C 335	A 301	E 439	B 338	D 528		−3.16
	k	296.68	292.28	468.28	330.88	552.88	388.20	
	$x-k$	38.32	8.72	−29.28	7.12	−24.88		
4	x	E 515	C 590	B 552	D 677	A 546		184.64
	k	564.08	563.08	489.68	626.48	636.68	576.00	
	$x-k$	−49.08	26.92	62.32	50.52	−90.68		
5	x	A 184	E 421	D 355	C 284	B 366		−69.36
	k	147.48	388.68	343.48	351.48	378.88	322.00	
	$x-k$	36.52	32.32	11.52	−67.48	−12.88		
Mean		286.00	364.60	378.00	407.00	521.20	391.36	
Deviation		−105.36	−26.76	−13.36	15.64	129.84		

Conditions	Mean	Deviation
A	322.20	−69.16
B	318.40	−72.96
C	405.20	13.84
D	426.20	34.84
E	484.80	93.44

Interaction SS $= \sum (x_{ij} - k_{ij})^2 = 44.391$

TABLE II.
$$y_{ij} = (k_{ij} - k...)^2/1000.$$

Animal	Period				
	1	2	3	4	5
1	B 302	D 2	C 0	A 24	E 519
2	D 351	B 469	A 397	E 1	C 7
3	C 90	A 98	E 59	B 37	D 261
4	E 298	C 295	B 97	D 553	A 602
5	A 595	E 0	D 23	C 16	B 2

S = Interaction Sum of Squares = 533,996

as the grand mean of the k's). Let h be the result of fitting to y in the same way as k was the fit to x (in the example, h is the fit of periods, animals and treatments to y). Then the sum of squares for non-additivity is

$$\frac{[\sum (y_{ij} - h_{ij})(x_{ij} - k_{ij})]^2}{\sum (y_{ij} - h_{ij})^2}$$

where, *in the numerator*, $(y_{ij} - h_{ij})$ can be replaced by y_{ij} without change in the value of the sum. (The choice in the numerator is a matter of arithmetic convenience. In the denominator, we *must* get at the sum of squares of the $y_{ij} - h_{ij}$, either directly, or by way of an analysis of variance.)

The original application to a balanced two-way design is another special case of the general procedure. There, however, the arithmetic is simplified if we use a seemingly quite different but numerically equivalent approach.

JOHN W. TUKEY

ERRATA

In Query 112, p. 568 of the December 1954 issue of Biometrics the following in Table III should be changed from

Differences of	$8 \geq 1, 2$	$8 \geq 1, 2$
established	$7 \geq 1$	$7 \geq 1$
sign at 5%	$6 \geq 1$	$6 \geq 1$

to

Differences of	$8 > 1, 2$	$8 > 1, 2$
established	$7 > 1$	$7 \geq 1$
sign at 5%	$6 > 1$	$6 > 1$

ABSTRACTS

Communication Prononcee A La Societe Francaise De Biometrie Le 24 Novembre 1954

302 A. HUET, D. SCHWARTZ, A. VESSEREAU. **Etude du Facteur "Sujet" et du Facteur "Vaccin" dans la Vaccination au B.C.G.**

Au cours de vaccinations collectives importantes effectuées par les soins du Centre International de l'Enfance, il a été possible de rechercher l'influence du facteur "ampoule de vaccin" en vaccinant plusieurs enfants avec chaque ampoule et en étudiant ensuite les points suivants: d'une part, on a examiné la répartition entre les ampoules des sujets demeurés non allergiques après la vaccination, d'autre part, pour les sujets allergiques, on a mesuré la dimension de l'induration consécutive au test tuberculinique, et recherché par analyse de la variance l'existence éventuelle d'un facteur "ampoule".

On a essayé en outre de caractériser un lot d'ampoules d'après les dimensions de l'induration mesurée sur les sujets; toutes les fois qu'on décèle l'existence du facteur "ampoule", les mesures correspondant à une même ampoule ne sont plus indépendantes; on est ainsi ramené à rechercher une valeur typique pour une collection de K objets mesurés chacun avec un nombre variable N_i de répétitions; il y a lieu de caractériser la collection par une moyenne pondérée des moyennes par objet.

Les auteurs ont proposé des formules donnant, en tenant compte du nombre variable d'enfants vaccinés par ampoule, des estimations de cette moyenne pondérée et de sa variance.

303 M. OLLAGNIER. **Utilisation des Fiches Perforees a 80 Colonnes pour l'Interpretation des Resultats des Experiences Agronomiques Factorielles.**

L'utilisation des cartés perforées à 80 colonnes, décrite par O Kempthorne pour les essais de type 2^5, permet à l'Institut de Recherches pour les Huiles et Oleagineux (I.R.H.O.) l'analyse rapide de séries d'essais factoriels (2^n, 3^n, $4 \times 4 \times 2$, $3 \times 3 \times 2$, $3 \times 2 \times 2$) pour lesquels un nombre élevé de facteurs est étudié. A chaque parcelle correspond une carte sur laquelle sont perforées d'une part les données expérimentales et d'autre part les participations positives ou négatives de la parcelle aux différents effets, chaque effet étant subdivisé en autant de fonctions linéaires que de degrés de liberté. Les interactions d'ordre élevé généralement négligeables sont utilisées pour estimer l'erreur. On évite ainsi tous les calculs classiques d'analyse de variance (sommes de carrés, terms de correction). Le procédé n'est financièrement rentable que si l'on traite un nombre suffisant d'essais et de facteurs par essai (10 à 15).

THE BIOMETRIC SOCIETY

Biometric Symposium in Brazil. The next international meeting of the Society will be a Biometric Symposium in Campinas, near São Paulo, Brazil. It has been scheduled for July 4–8, 1955, following the meetings in Rio de Janeiro of the Inter-American Statistical Institute on June 10–22 and of the International Statistical Institute on June 24–July 3, in which the Society has been invited to sponsor a program. Since The Biometric Society has the status of a Section in the International Union of Biological Sciences, the Symposium in Campinas also meets under the auspices of the Union. The travel funds that have been made available by the IUBS, by the National Science Foundation for United States citizens, and by other organizations for staff members are making it possible to arrange a varied and challenging program. The Symposium will consider the Role of Biometric Techniques in Biological Research, with sessions or papers on experiments with perennial crops, grazing and feeding experiments, biometrical genetics, population genetics, bioassay, sampling techniques and medical statistics. Local arrangements for the Symposium are being handled by Dr. C. C. Fraga, Instituto Agronomico, Campinas, Est. São Paulo, Brazil. The program and general plans are under the chairmanship of the President of The Biometric Society, Professor W. G. Cochran, Johns Hopkins University, Baltimore, Maryland, U.S.A. Anyone who plans to attend —the Symposium is open to all—is urged to write one of the above or to the Secretary of the Society, Box 1106, New Haven 4, Connecticut.

European Seminar in Biometry. Plans are progressing for a Seminar in Biometry next September under the sponsorship of the Italian Region. Lasting three weeks, it will provide courses, with laboratory exercises, on the biometrical aspects of the design and analysis of biological experiments. Through the courtesy of the Italian Government, the Seminar will meet in the famous Monastero Villa at Varenna on Lake Como. Twenty or more graduates from different branches of biology and related fields can be accommodated, and, thanks to a grant from the IUBS, expenses for each participant will be held to a minimum. All inquiries should be addressed to Dr. L. L. Cavalli-Sforza, Via Darwin 20, Milano, Italy, who is in charge of the project. We hope that similar Seminars can be continued in future years, rotating among different European countries.

WHO. Dr. Manuel Aycardo served as Observer for The Biometric Society at the Fifth Session of the Regional Committee for the Western Pacific of the World Health Organization in Manila, P.I., on September 10–16, 1954. Committee members representing 14 countries and

delegates from 22 international associations attended. One resolution passed by the Committee related to the appointment of a Regional Statistician. In his statement at the Session, Dr. Aycardo emphasized the need in health work to plan statistically and that failure to do this could make later evaluation of the work impossible.

Netherlands. Two biometric sessions, both at the University of Utrecht, were sponsored in 1954 by members of the Society in collaboration with two other Dutch biometrical clubs. On February 25, Professor J. Meertens and Dr. A. Drion gave papers on biometrical problems in genetics. In the meeting of October 27, lectures on the use of statistical methods in different branches of research were given by Th. J. D. Erlee (Uniformity trials in sugarcane), Dr. D. Dresden (Insecticides), Ir. Th. Ferrari (Multifactoranalysis), Ir. H. de Miranda (Organoleptic problems), A. A. van Soestbergen (Toxoplasmosis) and Ir. J. van Soest (Forestry problems). By courtesy of the Netherlands Statistical Society (Industries Section) members of the biometrical societies were invited to hear at Utrecht a paper read by Dr. Read (Manchester) on industrial experimentation.

ENAR. The Region met jointly with the Statistics Section of the American Public Health Association on October 13 in Buffalo, New York, during the annual meeting of the APHA. The Uses of Sampling in Public Health and Related Fields were considered in papers by M. Rosenstock on Application of sampling in the evaluation of health education material, by A. Bachrach on The application of sampling methods for calculating hospital stay, and by D. M. Schneider on Use of sampling techniques in the adjustment of uniform hospital rates.

The Biometric Society (ENAR) will meet with the America Institute of Biological Sciences at Michigan State College, East Lansing, on September 5–9, 1955. Titles and abstracts for contributed papers for The Biometric Society should be sent to Dr. Earl L. Green, Division of Biology and Medicine, U.S. Atomic Energy Commission, Washington 25, D.C., not later than May 15, 1955.

Region Francaise. La dernière réunion de la Societé a eu lieu le 24 Novembre au Laboratoire de Zoologie de L'Ecole Normale Supérieure, Paris. L'ordre du jour était le suivant: M. Ollagnier: L'utilization des fiches perforées pour l'interprétation des résultats des expériences factorielles agronomiques. Dr. A. Huet, D. Schwartz, A. Vessereau: Etude du facteur "sujet" et du facteur "vaccin" dans la vaccination au B.C.G.

Switzerland. At the November 27 meeting of the Swiss members of the Society, held in the Ophthalmological Clinic of the University of Geneva, the following papers were presented: La Biométrie en Suisse

by A. Linder, Expériences biométriques en Endocrinologie by R. Borth, Biometrical problems arising out of alcoholism by E. M. Jellinek, and L'organisation et la travail de la Division des Services d'Epidémiologie et de Statistiques sanitaires de l'Organisation mondiale de la Santé by Y. Biraud.

WNAR. During the Berkeley meetings of the American Association for the Advancement of Science, the Region co-sponsored three sessions on December 27-28, 1954, in collaboration with the Third Berkeley Symposium, the American Statistical Association, the Ecological Society of America and the Institute of Mathematical Statistics. The first, on Statistics in Biology and Genetics, featured papers on Struggle for existence by T. Park, J. Neyman and E. L. Scott, Some genetic problems in controlled populations by E. Dempster, and Some genetic problems in natural populations by J. F. Crow and M. Kimura. The Design of Experiments in Fisheries was the subject of the second program, with papers on Biological assumptions involved in estimating mortalities to downstream migrant salmon passing dams by C. O. Junge, Jr., Use of logbook data in the measurement of distribution and abundance of commercial fish stocks by M. B. Schaefer, and Some remarks on the design of a sampling program of a fishery for a measure of fishing intensity by T. M. Widrig. In the third session on Statistics in Medicine and Public Health, W. F. Taylor discussed Problems of contagion; C. L. Chiang and J. Yerushalmy, Statistical problems in medical diagnoses, and J. Cornfield, Some statistical problems arising from retrospective studies.

NOTES

Cooperative Graduate Summer Sessions in Statistics

The University of Florida, North Carolina State College, Virginia Polytechnic Institute and the Southern Regional Education Board are jointly sponsoring a series of cooperative summer sessions in statistics.

The first of these cooperative graduate summer sessions was held during the summer of 1954 at Virginia Polytechnic Institute. At this session there were 89 students from 26 states and the District of Columbia and from India, Finland, Canada, Australia, China, Hawaii and the Philippines. The following courses were offered: Engineering Statistics, Statistical Methods I, Statistical Theory I (Probability and Inference), Biostatistics, Quantitative Genetics, Rank Order Statistics, Multivariate Analysis, and Seminar on Recent Advances in Statistics. Classes ranged in size from 9 to 34, with an average of 20.

The second session will be held at the University of Florida from June 20 to July 29, 1955. A session is scheduled to be held at North Carolina State College in 1956, and another at Virginia Polytechnic Institute in 1957.

The summer sessions are designed to carry out a recommendation of the Southern Regional Education Board's Advisory Commission on Statistics, on which the three institutions initiating the program are represented. The sessions will be of particular interest to (1) research and professional workers who want intensive instruction in basic statistical concepts and who wish to learn modern statistical methodology; (2) teachers of elementary statistical courses who want some formal training in modern statistics; (3) prospective candidates for graduate degrees in statistics; (4) graduate students in other fields who desire supporting work in statistics; and (5) professional statisticians who wish to keep informed of advanced specialized theory and methods.

Each of the summer sessions will last six weeks and each course will carry approximately three semester hours of graduate credit. The program may be entered at any session, and consecutive courses will follow in successive summers. The summer work in statistics may be applied as residence credit at any one of the cooperating institutions, as well as certain other institutions, in partial fulfillment of the requirements for a master's degree. The catalog requirements for the degree must be met at the degree-granting institutions. Each doctoral candidate should consult with the institution from which he desires to obtain the degree regarding the applicability of the summer courses in statistics.

The faculty for the 1955 session at the University of Florida will include: Professor R. L. Anderson, North Carolina State College; Professor D. B. Duncan, University of Florida; Professor Boyd Harshbarger, Virginia Polytechnic Institute; Professor Carl E. Marshall, Oklahoma A. and M. College; Professor Herbert A. Meyer, University of Florida; Professor George E. Nicholson, Jr., University of North Carolina; Professor Phillip J. Rulon, Harvard University; Professor Walter L. Smith, University of North Carolina; and Professor Dudley E. South, University of Florida.

Courses to be offered this summer are: Statistical Methods I, Statistical Methods II (Design of Experiments), Statistical Theory I, Statistical Theory II (Inference and Least Squares), Advanced analysis I, Theory of Sampling, Theory of Statistical Inference, Mathematics for Statistics, Statistical Research in Education and Psychology and Seminar on Recent Advances in Statistics.

The total tuition fee will be $35 for the six-weeks term. The holder of a doctorate degree, upon acceptance, may register without the payment of any tuition fee. Living and other expenses at the University are reasonable. The University is in Gainesville, located in the rolling hills of North Central Florida, midway between the cooling breezes of the Gulf of Mexico and the Atlantic Ocean.

Inquiries should be addressed to:

> PROFESSOR HERBERT A. MEYER
> Statistical Laboratory
> University of Florida
> Gainesville, Florida

Summer Sessions at Berkeley, California

This year's program at the Statistical Laboratory of the University of California, Berkeley, California, consists of two sessions: June 20-July 30 and August 1-September 10, 1955. The faculty of the summer sessions will include Professor G. E. Bates of Mt. Holyoke College, South Hadley, Massachusetts; Professor J. Neyman, Professor Charles H. Kraft and Mr. Howard G. Tucker of the Statistical Laboratory, University of California.

The program includes undergraduate courses primarily meant for students transferring from other centers who would like to embark on advanced studies in Berkeley during the regular academic year. Professor Neyman will be available for consultations on work leading to higher degrees. There will be no graduate course program. However, graduate students may be interested in a series of lectures and seminars

to be given through July and August in connection with the second part of the Third Berkeley Symposium on Mathematical Statistics and Probability. The scholars who promised to participate in this event are: T. W. Anderson, Columbia University, M. S. Bartlett, University of Manchester, J. Berkson, Mayo Clinic, David Blackwell, Howard University and University of California, A. J. L. Blanc-Lapierre, Université d'Alger, J. Doob, University of Illinois, W. Feller, Princeton University, R. Fortét, Institut Henri Poincaré, A. Girshick, Stanford University, J. M. Hammersley, Oxford University, J. L. Hodges, Jr., · University of California, W. Hoeffding, University of North Carolina, Lucien LeCam, University of California, Erich L. Lehmann, University of California, P. Lévy, l'École Polytechique, H. Robbins, Columbia University, Herman Rubin, Stanford University, and C. M. Stein, Stanford University.

Summer Offerings in Statistics at Iowa State College

The Department of Statistics at Iowa State College will offer a course in decision theory at the advanced graduate level during the first half of the 1955 summer quarter. The course will be taught by Dr. S. L. Isaacson. Members of the graduate faculty in statistics will be available during most of the summer for consultation on graduate research (Stat. 699) and for special problems courses (Stat. 599).

Other offerings for the two six-week sessions (June 13-July 20 and July 20-August 26) of the summer quarter are designed mainly for the graduate minor in statistics and for the beginning graduate major in statistics who wish to satisfy prerequisite requirements for more advanced courses. These additional offerings include Stat. 401 and 402, "Statistical Methods for Research Workers," offered in sequence; the sequence, Stat. 447 and 448, "Statistical Theory for Research Workers;" Stat. 411, "Experimental Designs for Research Workers;" and Stat. 421, "Survey Designs for Research Workers." Students may register for one or both summer sessions. For additional information, write to: T. A. Bancroft, Director, The Statistical Laboratory, Iowa State College, Ames, Iowa.

*Joint Meeting of the Institute of Mathematical Statistics
and The Biometric Society (ENAR)*

FRIDAY, APRIL 22

8:30 A.M.   Invited Speakers

Chairman: ˙Professor H. Fairfield Smith, North Carolina State College

"Life Testing in the Discrete Case"*—Franklin S. McFeely and John E. Freund, Virginia Polytechnic Institute

"The Components of Variance and the Correlation Between Relatives in Symmetrical Random Mating Populations"—Ted Horner, Iowa State College

"Tests of Hypotheses When the Decision is Based on Several Criteria"* (Preliminary Report)—Irwin Miller and John E. Freund, Virginia Polytechnic Institute

"Power Function of Procedures for Some Components of Variance Models"—Helen Bozivich, Iowa State College

"Preference Patterns for Decisions on Means"*—R. Lowell Wine and John E. Freund, Virginia Polytechnic Institute

*Research sponsored by the Office of the Ordnance, U. S. Army

10:30 A.M.   Probability Theory

Chairman: Dr. Eugene Lukacs, Office of Navy Research

Speakers: D. Austin—Syracuse University
J. Blackman—Syracuse University
Cyrus Derman—Syracuse University

2:00 P.M.   Multivariate Analysis

Chairman: Dr. Harold Hotelling, University of North Carolina

Speakers: T. W. Anderson—Columbia University
W. G. Howe—Oak Ridge Institute of Nuclear Studies
H. C. Sweeny, Virginia Polytechnic Institute

4:00 P.M.   Contributed Papers

Chairman: Dr. George E. Nicholson, Jr., University of North Carolina

*Abstracts received prior to March 1, 1955.

SATURDAY, APRIL 23

8:30 A.M.   Relation Between Smoking and Mortality From Lung
            Cancer

            Chairman: Dr. B. G. Greenberg, University of North
            Carolina

            Speakers: William Haenszel, National Cancer Institute
                      Jerome Cornfield, National Institutes of Health
                      Joseph Berkson, Mayo Clinic and University of
                      Minnesota

            Discussants: Boyd Harshbarger, Virginia Polytechnic
                      Institute
                      Daniel Horn, American Cancer Society

10:30 A.M. Contributed Papers*

            Chairman: Dr. R. L. Anderson, North Carolina State College
            1. Information and Distance Applied to Discriminant
               Analysis Between Two Normal Populations—Samuel W.
               Greenhouse, National Institute of Mental Health.
            2. Appropriate Scores in Bio-assays Using Death-Times
               and Survivor Symptoms—Johannes Ipsen, Institute of
               Laboratories and Harvard School of Public Health.
            3. A Comparison of Random and Non-Random Plot Selec-
               tion—Daniel G. Horvitz and Jack Fleischer, North
               Carolina State College.

*Abstracts received prior to March 1, 1955.

RULES OF THUMB FOR DETERMINING EXPECTATIONS OF MEAN SQUARES IN ANALYSIS OF VARIANCE*

E. F. SCHULTZ, JR.

Alabama Polytechnic Institute and
Institute of Statistics, North Carolina State College

INTRODUCTION

Exact procedures for determining the expected values of sample mean squares in terms of population parameters are adequately described in a number of places in statistical literature (1, 3, 7)†. For simple designs with few classifications the processes can be gone through quickly, and with practice, the expectations of such mean squares can be written by inspection. However, when a design involves several classifications, and particularly when the classifications are a mixture of random and fixed variates, the processes become complex and tedious.

The purpose of this paper is to illustrate a set of simple rules which reduces the processes of determining the expectations of the mean squares of even complex analyses to practically the equivalent of determination by inspection. These rules are sufficiently general to cover all complexities of classification, provided the sums or means at each level of summarization are composed of equal numbers of observations and, in the case of random variates, are drawn from infinite populations.

With respect to fixed and random effects two population models are of common occurrence (1, 5, 6):

(1) every variate random so that all components are random except the general mean (Eisenhart's Model II)
(2) a mixture of random and fixed variates known oftentimes as the mixed model.

Since random variates have a probability distribution but fixed effects do not, it is necessary to determine for each factor under investigation whether its effects are to be regarded as fixed or random (1).

In general, if all the treatments (or classifications) about which inferences are to be made are included in an experiment (or survey) the treatments or classifications are regarded as fixed. Since it would be

---

†Numbers in parentheses refer to references cited.

most unusual to make inferences about treatments or classifications not included in an experiment (except by transformation and interpolation of quantitative classifications) it follows that the treatments or classifications studied in an experiment are the only treatments about which inferences are planned (i.e., are the complete population of treatments so far as a particular experiment is concerned) and therefore treatments are customarily regarded as fixed.

If on the other hand it is wished to make inferences about an overall mean effect from a sample only of all the effects such as, perhaps, the average yield of inbred lines of corn from the observed performance of only a few lines, then the effects are regarded as random.

The sampling or experimental design and procedures (which must be known for analysis) are also helpful in determining whether effects are to be regarded as fixed or random.

<div align="center">THE RULES</div>

For Both Models

RULE 1.   Decide for each variate (sampling level or factor) whether it is to be regarded as fixed or random and assign it a letter to be used both as a designating symbol and as a coefficient indicating the number of such individuals. List the sources of variation in the analysis of variance, completely identifying each source by means of the selected symbols.

It is helpful in naming the sources of variation and components, and in preventing omissions of components, if sources are listed in hierarchal order.   Hierarchal is used in its broader sense to include hierarchy involving cross classified variates as occurs in the split plot design.

RULE 2.   List in the expectation of each mean square the component due directly to that particular source. Completely identify the component by using as subscripts all of the symbols necessary to completely identify or describe the source; in which case all of the remaining symbols become coefficients of the component.   This procedure completely identifies the totality of components which must be considered.   List as other components in the expectation of a particular mean square all other components whose identifying subscripts contain all of the symbols necessary to completely describe the source of the mean square under consideration.

It is helpful if the order of the subscripts is such that the first symbols following σ^2 describe the origin of the variation while the remainder (enclosed in parentheses) indicate the position in the hierarchy at which the component arises.   The subscripts describing the origin of the

variation will, for purposes of distinction, be referred to as "essential" or "truly descriptive". If the suggested procedure of ordering subscripts is followed (as it is in this paper) we may define the "essential" or "truly descriptive" subscripts in a mechanical manner as those immediately following σ^2 and not enclosed by parentheses.

For the Mixed Model

If there are fixed effects (either one or more) then Rules 1 and 2 still hold by virtue of adding Rule 3 specifying certain deletions from expectations obtained by Rules 1 and 2.

RULE 3. To determine which components should be deleted consider each component in the following manner. Among the "essential" or "truly descriptive" subscripts of the component under consideration ignore or delete from consideration those one or more subscript symbols which are necessary to describe the source of variation in which the component is listed. *If any of the remaining "essential" subscripts specifies a fixed effect, delete the component from the expectation.*

The necessity for Rule 3 arises from the fact that in the case of a fixed effect the total population has been included and there is no component of uncertainty in the estimate due to having *sampled* the population. If the method of sampling leads to cross classification of a fixed effect with a random variate then the resulting interaction gives rise to a component which is "random in one direction only", i.e., such a component *does exist* as a part of the expectation of the mean square of the fixed effect (since measured over the random variate) but *does not exist* as a part of the expectation of the random variate (since measured over the fixed effect) (1).

For purposes of distinction a component due directly to a fixed effect is denoted by θ^2.

<div align="center">EXAMPLES</div>

An Example with Simple Sampling and Subsampling, All Variates Random

Suppose, in order to estimate the firmness of peaches in a certain location during a particular season, one may have made duplicate determinations of the firmness of peaches chosen in the following manner: a definite number of peaches chosen at random from each tree of a sample of trees in the location.

Following Rule 1 we list the sources of variation as in the first column of Table 1. It is convenient to designate trees by t which, when used as a coefficient, also designates the number of trees. Since the trees

are only a random sample of the trees producing the peaches whose firmness we wish to estimate, we may correctly decide that trees are random.

<div align="center">TABLE 1</div>

Structural Analysis and E(M.S.) for a Sampling Scheme Investigating Fruit Firmness by Means of d Duplicate Determinations on Each of f Fruit from Each of t Trees, all Components Random Except the Mean.

Source of Variation	d.f.	E(M.S.)
Total	$dft - 1$	
Trees (T)	$t - 1$	$\sigma^2_{d(f)(t)} + d\sigma^2_{f(t)} + df\sigma^2_t$
Fruits (F) in T	$(f - 1)t$	$\sigma^2_{d(f)(t)} + d\sigma^2_{f(t)}$
Detns. (D) in F in T	$(d - 1)ft$	$\sigma^2_{d(f)(t)}$

Fruit may be designated by f which, when used as a coefficient, also designates the number of fruit *per tree*. Since the individual fruit were chosen by random means, they are properly regarded as random samples of the fruit on the trees from which they were harvested.

The duplicate determinations made on each fruit are designated by d which, when used as a coefficient, also designates the number of determinations *per fruit*. Duplicates can hardly be regarded otherwise than as representing random effects.

We see now that the model with all components random except the general mean is appropriate.

Following Rule 2 we list for each source of variation a component due directly to that source. For each mean square this is the component listed last. For the last listed source of variation, that of the ultimate units of the experiment, we find the component to be $\sigma^2_{d(f)(t)}$ which is the expected mean square of that source, Determinations in Fruit in Trees. It sometimes happens that the basic unit of variation represents two or more components, but if so, they are confounded and are treated as a single component.

Advancing to Fruit in Trees it is easily verified that the subscripts in $\sigma^2_{d(f)(t)}$ contain f and t, the symbols necessary to fully describe the source, Fruit in Trees, hence $\sigma^2_{d(f)(t)}$ is a part of the expectation of the mean square of this particular source. There is also the component due directly to the source, in this case $\sigma^2_{f(t)}$. Since this component requires only f and t for designation, the remaining symbols, only d in this case, appear as coefficients giving $d\sigma^2_{f(t)}$. The expectation of M.S.$_{F\,in\,T}$ is $\sigma^2_{d(f)(t)} + d\sigma^2_{f(t)}$ as shown in Table 1.

Advancing now to consideration of the expectation of M.S.$_T$ we

find that $\sigma^2_{d(f)(t)}$ contains t, so that $\sigma^2_{d(f)(t)}$ is part of the expectation of Trees. Also $\sigma^2_{f(t)}$ contains t so that the component due directly to the Fruits in Trees ($\sigma^2_{f(t)}$ with coefficient d) is also a part of the expectation of Trees. There is also a component due directly to Trees, σ^2_t, with the remaining symbols as coefficients yielding $df\sigma^2_t$. The expectation of M.S.$_T$ is then $\sigma^2_{d(f)(t)} + d\sigma^2_{f(t)} + df\sigma^2_t$ as shown in Table 1.

An Example with Both Cross Classification and Sampling, All Variates Random

Suppose now, that in order to take account of the day to day variability which may exist, we repeat the sampling procedure on the same trees on each of several days not chosen for any characteristic.

Following Rule 1 we assign q to indicate days when used as a subscript and to indicate the number of days when used as a coefficient. The days are to be regarded as having random effects since they were not chosen to represent any special characteristic of days and no inferences about the effects of various kinds of days are contemplated.

We may observe that again we have the model with all components random except the general mean. At some levels we have again used simple random sampling (fruits and duplicate determinations). As regards days and trees however, while each was selected in a random fashion, observations were repeated on the *same* trees on the different days. This leads to cross classification of the observations and one of the sources of variation will now be the result of interaction or discrepancy.

The sources of variation in this experiment are shown in the first column of Table 2.

TABLE 2

Structural Analysis and E(M.S.) for a Sampling Scheme Investigating Fruit Firmness by Means of d Duplicate Determinations on Each of f Fruits from Each of t Trees, the Whole Repeated on the Same Trees on q Days, All Components Random Except the Mean.

Source of Variation	d.f.	E(M.S.)
Total	$dfqt - 1$	
Trees (T)	$t - 1$	$\sigma^2_{d(f)(qt)} + d\sigma^2_{f(qt)} + df\sigma^2_{qt} + dfq\sigma^2_t$
Days (Q)	$q - 1$	$\sigma^2_{d(f)(qt)} + d\sigma^2_{f(qt)} + df\sigma^2_{qt} + dft\sigma^2_q$
$Q \times T$	$(q - 1)(t - 1)$	$\sigma^2_{d(f)(qt)} + d\sigma^2_{f(qt)} + df\sigma^2_{qt}$
Fruits (F) in $Q \times T$	$(f - 1)qt$	$\sigma^2_{d(f)(qt)} + d\sigma^2_{f(qt)}$
Detns. (D) in F in $Q \times T$	$(d - 1)fqt$	$\sigma^2_{d(f)(qt)}$

Listing for each source of variation the component due to that source we find opposite M.S.$_{DinFinQ\times T}$, the source of unit variance, its expectation $\sigma^2_{d(f)(qt)}$.

Considering M.S.$_{FinQ\times T}$ it is plain that the subscripts of $\sigma^2_{d(f)(qt)}$ contain f, q, and t, the symbols necessary to identify the source under consideration so the component $\sigma^2_{d(f)(qt)}$ is a component of the expected value of M.S.$_{FinQ\times T}$. This component together with the component due directly to the source, $\sigma^2_{f(qt)}$ with coefficient d, comprise the expectation of M.S.$_{FinQ\times T}$.

The procedure is followed until we find the expectation of M.S.$_T$ to be

$$\sigma^2_{d(f)(qt)} + d\sigma^2_{f(qt)} + df\sigma^2_{qt} + dfq\sigma^2_t .$$

A More Complex Example with Both Cross Classification and Sampling, All Variates Random

Actually such an experiment as described in the previous example might be repeated at a number of locations in order to obtain an estimate for the region rather than a particular location (Table 3). It might

TABLE 3

Structural Analysis and E(M.S.) for a Sampling Scheme Investigating Fruit Firmness by Means of d Duplicate Determinations on Each of f Random Fruit from Each of t Random Trees in Each of l Random Locations, the Whole System Repeated on the Same Trees on Each of q Random Days.

Source of Variation	d.f.	$\sigma^2_{d(f)(qt)(l)}$	$d\sigma^2_{f(qt)(l)}$	$df\sigma^2_{qt(l)}$	$dft\sigma^2_{ql}$	$dfl\sigma^2_q$	$dfq\sigma^2_{t(l)}$	$dfql\sigma^2_i$
Total	$dfqtl - 1$							
Locations (L)	$l - 1$	x	x	x	x		x	x
Trees (T) in L	$(t - 1)l$	x	x	x			x	
Days (Q)	$q - 1$	x	x	x	x	x		
$Q \times L$	$(q - 1)(l - 1)$	x	x	x	x			
$Q \times T$ in L	$(q - 1)(t - 1)l$	x	x	x				
Fruits (F) in $Q \times T$ in L	$(f - 1)qtl$	x	x					
Detns. (D) in F in $Q \times T$ in L	$(d - 1)fqtl$	x						

also be that, though the days were randomly chosen, the work was so coordinated that the observations were made on the *same* days at the different locations.

Following Rule 1 we assign the symbol l to locations and decide, since the locations were chosen only to represent the region, that locations are to be regarded as a random variate.

Further application of the rules leads to the expectations in Table 3. Instead of writing out each component with its necessary list of coefficients and subscripts each time it occurs in Table 3, there is provided for each component a column which is merely checked if the component is a part of the expectation of a mean square under consideration. This example demonstrates that, even with a complex experiment, application of the proposed rules leads to the correct expectations. It will be used later to illustrate Rule 3.

An Example of Cross Classification, Fixed Effects with One Random Variate

It is entirely possible that one's primary aim in investigating peaches could have been to determine whether different pruning methods applied to peach trees affect the firmness of the fruit differently. In this case one might have selected several *blocks of trees*, which because of their appearance and contiguity were judged to be similar trees, and have allotted the pruning treatments one per tree to the several trees of a block, repeating the procedure in each block. The plan of selecting f fruit from each tree and making d determinations on each fruit might well have been continued. Suppose we have data at hand collected by such a procedure and that there are results for one day only.

Following Rule 1 we would conclude that determinations and fruit are still random. Trees also are still random but they have been replaced by *blocks of trees*, or *replications*, which give observations that are cross classifiable with respect to prunings. The pruning, however, is entirely at the disposal of the experimenter. He will choose to prune in certain fashions, and he will draw inferences about the effects of pruning in these certain fashions, but in no other. For purposes of consideration, then, the entire population of pruning methods is represented in the experiment. As a consequence there is no variability due to sampling the population of pruning methods and we consider the effects of prunings to be fixed (or constant).

We have then p fixed prunings on single trees in each of r random replications, with f random fruit per tree, and d random duplicate determinations per fruit.

Application of Rules 1 and 2 leads to the components listed in Table 4.

TABLE 4

Structural Analysis and E(M.S.) for a Sampling Scheme Investigating Fruit Firmness of p Fixed Prunings Imposed on Single Trees in Each of r Random Replications with f Random Fruit per Tree and d Random Determinations per Fruit.

Source of Variation	d.f.	E(M.S.)*
Total	$dfrp - 1$	
Replications (R)	$r - 1$	$\sigma^2_{d(f)(pr)} + d\sigma^2_{f(pr)} + \underline{df\sigma^2_{pr}} + dfp\sigma^2_r$
Prunings (P)	$p - 1$	$\sigma^2_{d(f)(pr)} + d\sigma^2_{f(pr)} + \overline{df\sigma^2_{pr}} + dfr\theta^2_p$
$P \times R$	$(p - 1)(r - 1)$	$\sigma^2_{d(f)(pr)} + d\sigma^2_{f(pr)} + df\sigma^2_{pr}$
Fruits (F) in $P \times R$	$(f - 1)pr$	$\sigma^2_{d(f)(pr)} + d\sigma^2_{f(pr)}$
Detns. (D) in F in $P \times R$	$(d - 1)fpr$	$\sigma^2_{d(f)(pr)}$

*Underscored components do not exist under the conditions assumed.

Applying Rule 3 to component $df\sigma^2_{pr}$ in the expectation of the mean square for replications, $E(\text{M.S.}_R)$, we find that we are required to ignore or delete or cancel from consideration, "essential" or "truly descriptive" subscript r (immediately following σ^2 and not enclosed in parentheses) because the symbol r is required in the description of the source. This leaves only subscript p. Since p, a remaining "essential" subscript, represents a fixed effect the component is deleted from the expectation. The deletion is indicated in Table 4 by underscoring $df\sigma^2_{pr}$ so that $E(\text{M.S.}_R)$ is $\sigma^2_{d(f)(pr)} + d\sigma^2_{f(pr)} + dfp\sigma^2_r$. This is the *only* component deleted from Table 4 by application of Rule 3.

A Complex Example of Cross Classification, Two Sets of Fixed Effects which Cross Classify with Two Random Variates which Cross Classify

In actuality the investigator might simultaneously investigate the effect of pruning on firmness of both ripe and green peaches and, as in our second example, he might also investigate whether there were day to day variations in the effects.

There would then be pm combinations of p fixed prunings with m fixed maturities investigated on single trees in r random replications repeated on the *same* trees on each of q random days with f fruit being taken at random from each tree each day with d duplicate determinations of firmness being made on each fruit.

We proceed again by Rules 1 and 2 laid down for the case of all variates random with the idea that we will later use Rule 3 to strike out such components as do not exist because of the different behavior of components when the model includes fixed effects. We have then Table 5.

TABLE 5

Structural Analysis and E(M.S.) for a Sampling Scheme Investigating Fruit Firmness of *m* Fixed Maturities on Each of *p* Fixed Prunings Imposed on Single Trees in Each of *r* Random Replications with Observations Repeated on the Same Trees on Each of *q* Random Days by Means of *d* Duplicate Determinations on Each of *f* Random Fruit of Each Maturity from Each Tree.

E(M.S.)*

Source of Variation	d.f.	$\sigma^2_{d(fmpqr)}$	$d\sigma^2_{f(mpqr)}$	$df\sigma^2_{mpqr}$	$dfr\sigma^2_{mpq}$	$dfq\sigma^2_{mpr}$	$dfqr\theta^2_{mp}$	$dfp\sigma^2_{mqr}$	$dfpr\sigma^2_{mq}$	$dfpq\sigma^2_{mr}$	$dfpqr\theta^2_m$	$dfm\sigma^2_{pqr}$	$dfmr\sigma^2_{pq}$	$dfmq\sigma^2_{pr}$	$dfmqr\theta^2_p$	$dfmp\sigma^2_{qr}$	$dfmpr\sigma^2_q$	$dfmpq\sigma^2_r$
Total	$dfmpqr - 1$																	
Replications (R)	$r - 1$	×	×	*×*		*×*		*×*		*×*		*×*		*×*		×		×
Days (Q)	$q - 1$	×	×	*×*	*×*			*×*	*×*			*×*	*×*			×	×	
$Q \times R$	$(q-1)(r-1)$	×	×	*×*				*×*				*×*				×		
Prunings (P)	$p - 1$	×	×	*×*	*×*	*×*	*×*					×	×	×	×			
$P \times R$	$(p-1)(r-1)$	×	×	*×*		*×*						×		×				
$P \times Q$	$(p-1)(q-1)$	×	×	*×*	*×*							×	×					
$P \times Q \times R$	$(p-1)(q-1)(r-1)$	×	×	*×*								×						
Maturities (M)	$m - 1$	×	×	*×*	*×*	*×*	*×*	×	×	×	×							
$M \times R$	$(m-1)(r-1)$	×	×	*×*		*×*		×		×								
$M \times Q$	$(m-1)(q-1)$	×	×	*×*	*×*			×	×									
$M \times Q \times R$	$(m-1)(q-1)(r-1)$	×	×	*×*				×										
$M \times P$	$(m-1)(p-1)$	×	×	×	×	×	×											
$M \times P \times R$	$(m-1)(p-1)(r-1)$	×	×	×		×												
$M \times P \times Q$	$(m-1)(p-1)(q-1)$	×	×	×	×													
$M \times P \times Q \times R$	$(m-1)(p-1)(q-1)(r-1)$	×	×	×														
Fruits (F) in $M \times P \times Q \times R$	$(f-1)mpqr$	×	×															
Detns. (D) in F in $M \times P \times Q \times R$	$(d-1)fmpqr$	×																

*Italicised components do not exist under the conditions assumed.

For a specific example of the operation of Rule 3 consider in Table 5 the expectation of Prunings mean square, $E(\text{M.S.}_P)$. Starting with components due to smaller units in the first 2 columns we note that the "essential" subscripts of $\sigma^2_{d(f)(mpqr)}$ and $d\sigma^2_{f(mpqr)}$ include only subscripts representing random variates so that the conclusion regarding presence or absence of these components will *not* be affected by the application of Rule 3.

In the third column we find a component due to interaction $df\sigma^2_{mpqr}$ with 4 "essential" subscripts. Deleting p the symbol necessary to describe Prunings we have remaining m, q, and r. Since m, one of the remaining "essential" subscripts, represents a fixed effect this component, which would exist as a part of the expectation of Prunings if all variates were random, is *not* a part of the expectation under the assumption that maturities are fixed. In the next column we find the component $dfr\sigma^2_{mpq}$, whose "essential" subscripts contain m and q after deleting p. Since m represents a fixed effect this component does not exist in the expectation of Prunings. The presence of m in the "essential" subscripts of component $dfq\sigma^2_{mpr}$ and component $dfqr\theta^2_{mp}$ also precludes these components being a part of $E(\text{M.S.}_P)$. The next three components to be considered are $dfm\sigma^2_{pqr}$, $dfmr\sigma^2_{pq}$, and $dfmq\sigma^2_{pr}$. In each case, after deleting p, the subscript necessary to describe Prunings, the remaining "essential" subscripts represent only random variates, qr, q, and r respectively, so that these components *are* a part of $E(\text{M.S.}_P)$. It should hardly be necessary to remark that $dfmqr\theta^2_p$ is necessarily a part of $E(\text{M.S.}_P)$.

A MORE DIRECT PROCEDURE APPLICABLE TO ISOLATED MEAN SQUARES

Now that the rules of thumb have been enumerated and illustrated it may be meaningful to state the composition of an expected mean square more directly.

The expectation of any mean square contains, in addition to a component due directly to the source under consideration, all those components whose subscript symbols include the *set* of symbols necessary to completely describe the source, provided there are *only random variates represented in the "essential" subscripts after cancelling* those symbols necessary to describe the source of variation under consideration.

Examples

In the case illustrated in Table 4 the expected mean square for Prunings contains, in addition to the component due directly to Prun-

ings, two components due to the two random sampling variates, Fruit and Determinations, and a single component representing interaction or discrepance resulting from the cross classification of Prunings with a single random variate, Replications, thus:

$$E(\text{M.S.}_P) = \sigma^2_{d(f)(pr)} + d\sigma^2_{f(pr)} + df\sigma^2_{pr} + dfr\theta^2_p .$$

In the case illustrated in Table 5 the expectation of Prunings mean square contains, in addition to the component due directly to Prunings, the two components due to the two sampling variates, Fruit and Determinations, plus three components representing interactions of Prunings with the three forms of variability, Replications (R), Days (Q), and $Q \times R$, resulting from the cross classification of the two random variates Replications and Days, thus:

$$\begin{aligned} E(\text{M.S.}_P) = \ & \sigma^2_{d(f)(mpqr)} + d\sigma^2_{f(mpqr)} + dfm\sigma^2_{pqr} \\ & + dfmr\sigma^2_{pq} + dfmq\sigma^2_{pr} + dfmqr\theta^2_p . \end{aligned}$$

Should it have been the case that maturities were also regarded as random, then there would have been three random variates expressed in seven different forms $(R, Q, QR, M, MR, MQ, \text{and } MQR)$ so that $E(\text{M.S.}_P)$ would include, in addition to the component due directly to Prunings and the two components due to the sampling variates, seven components resulting from interaction or discrepance.

$$\begin{aligned} E(\text{M.S.}_P) = \ & \sigma^2_{d(f)(mpqr)} + d\sigma^2_{f(mpqr)} + df\sigma^2_{mpqr} \\ & + dfr\sigma^2_{mpq} + dfq\sigma^2_{mpr} + dfqr\sigma^2_{mp} \\ & + dfm\sigma^2_{pqr} + dfmr\sigma^2_{pq} + dfmq\sigma^2_{pr} + dfmqr\theta^2_p . \end{aligned}$$

That it is necessary to define the "essential" or "truly descriptive" subscripts, as opposed to those which merely denote the position in the hierarchy at which a component arises, may be shown by considering again the case illustrated in Table 3 but assuming now that Locations represent fixed effects.

When Rule 3 is properly applied under this assumption, the only deletion is component $dft\sigma^2_{ql}$ from the expectation of Days, $E(\text{M.S.}_Q)$. But should one forget to distinguish between the "essential" subscripts and subscripts in general, remembering only that Locations represent fixed effects, then, considering the source Days, and ignoring or cancelling the subscript q necessary to describe the source, one would find l remaining in each component of Days excepting σ^2_q , thus indicating that all random components should be deleted. This is obviously incorrect.

In Table 3 it is also interesting to observe the deletions due to regarding Days as fixed. In this case the component $df\sigma^2_{at(l)}$ is deleted from the expectation of Trees (T) in L and the two components $df\sigma^2_{at(l)}$ and $dft\sigma^2_{al}$ are deleted from the expectation of Locations.

<div align="center">SPECIAL SITUATIONS</div>

The Basic Unit of Variation is the Result of Interaction or Discrepance

A special case that is frequently met is an experiment conducted as that illustrated in Table 5 except that the firmness determination is made by *one* determination only on *one* fruit only from each tree on each date. In this case the basic component would be described as σ^2_{mpqr} , a component due to interaction. It must be recognized however that this estimate of σ^2_{mpqr} is confounded with components due to sampling variates such as fruit and determinations, and perhaps even others. Since it is unknown in this case whether σ^2_{mpqr} is large or small relative to the other components with which it is confounded the manner of treating σ^2_{mpqr} , the basic unit of variation, is uncertain. It would seem wise, in most cases at least, to treat this basic unit of variation as a component due to a single random sampling variate rather than an interaction, in which case it would be unaffected by Rule 3 concerning deletions.

The Factorial with a Single Error Term

If one is considering a factorial experiment of the type having p fixed prunings with f fixed fertilizers, the pf treatment combinations having been allotted at random to single trees in each of r replications, then the structural analysis usually is of the form following with the idea that "Pruning-Fertilizer Combinations" will be broken into an orthogonal set of comparisons for testing against a single error term.

Source	d.f.
Total	$rpf - 1$
Replications (R)	$r - 1$
Pruning—Fertilizer Combinations (C)	$pf - 1$
Error	$(pf - 1)(r - 1)$

To consider in this case that both Prunings and Fertilizers are separate fixed effects and to blindly isolate the interaction of each of these (and their joint effect) with replications according to the foregoing rules will lead to a separate error term with different expectation for each effect considered. To reconcile this circumstance with the originally proposed structural analysis, one has only to remember that one of

the basic assumptions of this type of analysis is that the errors are homogeneous and that, therefore, such components as σ_{pr}^2 for Pruning $\times$ Replication, σ_{fr}^2 for Fertilizer $\times$ Replication, and σ_{pfr}^2 for Pruning $\times$ Fertilizer $\times$ Replication are really estimates of the same component and therefore the three mean squares should be pooled as, say, σ_{cr}^2 for Pruning-Fertilizer Combinations $\times$ Replication.

Another matter exists which should be called to the reader's attention. When treatments are tried over two or more random variates which cross classify, none of the existing mean squares of the analysis of variance has the correct expectation to serve as error for testing the significances of differences among treatments. This situation exists in Tables 3 and 5. Error terms of the correct expectation may be constructed (1, 2, 8, 9).

REFERENCES

1. Anderson, R. L., and Bancroft, T. A., *Statistical Theory in Research*, McGraw-Hill (1952).
2. Cochran, W. G., "Testing a Linear Relation Among Variances", *Biometrics* V7: 17–32 (Mar. 1951).
3. Comstock, R. E., The Expectations of Mean Squares, *N. C. Institute of Statistics Mimeograph Series No. 76*, (1953).
4. Crump, S. L., "The Estimation of Variance Components in Analysis of Variance", *Biometrics* V2: 7–11 (1946).
5. Crump, S. L., "The Present Status of Variance Component Analysis", *Biometrics* V7: 1–16 (Mar. 1951).
6. Eisenhart, Churchill, "The Assumptions Underlying the Analysis of Variance", *Biometrics* V3: 1–21 (Mar. 1947).
7. Kempthorne, Oscar, *Design and Analysis of Experiments*, John Wiley and Sons (1952).
8. Satterthwaite, F. E., "An Approximate Distribution of Estimates of Variance Components", *Biometrics* V2: 110–114 (1946).
9. Schultz, E. F. Jr., "Query," *Biometrics* V10: 407–411 (1954).

VARIANCE COMPONENTS WITH REFERENCE TO GENETIC POPULATION PARAMETERS*

DOROTHY C. LOWRY

University of California

The nature of variability in a population observed at a given time with respect to a particular metric trait is of great interest and importance among animal and plant breeders. Their work depends heavily upon the ability to design breeding experiments and to take advantage of statistical techniques which will enable them successfully to apportion differences in such a trait to the various broad causal factors operating upon the individuals constituting the population. This they must accomplish with sufficient accuracy to describe to some extent the genetic and environmental complex affecting the trait and to predict breeding results. It has been shown, particularly by the work of Fisher, Haldane and Wright, that for various quantitative traits the system of genes involved does have average properties which are measurable and the analysis of variance has proved to be a powerful tool in the estimation of such parameters. This paper is presented as a review of some of the applications of variance components in statistical genetics and of some statistical problems commonly encountered in their use in this field.

The situation frequently to be met in quantitative genetics is as follows: we have a set of data arranged in a particular type of classification and described by a linear function of effects of various classes and subclasses. Generally this model is that which Eisenhart (1947) has called Model II, in which all elements except μ are regarded as random variables, although it may frequently be what he called the Mixed Model, in which certain of the effects are regarded as fixed rather than as random variables. The first step then is the estimation of the variances of these random variables and the second step the linear combination of certain of these estimates to provide further estimates of the parameters of heredity, by which I mean any of the parameters, genetic and environmental, describing the variability of the quantitative trait.

Weinberg (1910) showed that the correlation between parent and offspring is $1/2 \ \sigma_G^2/\sigma_T^2$ in a random breeding population, where σ_G^2 is

---

the genetic variance and σ_T^2 is the total variance, if it can be assumed that the genetic component is due entirely to autosomal factors with effects which are additive. Fisher (1918) examined the correlations between relatives with respect to a metric trait to be expected under the Mendelian hypothesis, that is, under the assumption that such traits are determined by a large number of segregating genes distributed among the chromosomes. He considered both random and assortative types of mating and demonstrated the effects upon these correlations of non-additive gene action of two types:

1) dominance deviations, which pertain to single pairs of allelic genes. When such deviations exist the values of the three genotypes AA, Aa and aa, each averaged over the array of environments to which the population is subjected, are a, d and $- a$, respectively, where d may have any value from $- a$ to a and even values outside this range. In the case of no dominance, the three genotypes would be represented by, say, $b + c$, b and $b - c$, respectively; that is the heterozygote would be midway between the two homozygotes in value.
2) epistatic deviations which arise from interactions between non-allelic pairs of genes.

Thus Fisher divided the genetic variance of a breeding population into the additively genetic variance, the variance due to dominance deviations from the additive scheme and the epistatic variance and showed the decrease to be expected in the correlations between individuals of various relationships due to the operation of dominance and epistasis. The extensive work of Wright (1917, 1918, 1920, 1921, 1935) on the correlations between any relatives as well as extensions and applications by a number of people working in the field of quantitative genetics (Mather, Lush, Lerner, et al.) enable us to partition the phenotypic variance of a population into an additively genetic portion and an environmental proportion under a number of assumptions of which the most important are:

1) Gene differences have strictly additive average effects over the array of environments of the population.
2) No correlation exists between the average value of a genotype and its environmental variance.
3) Hereditary and environmental factors are not correlated in occurrence.
4) Random mating obtains, or a mating plan in which the non-randomness can be expressed quantitatively.

The breeding experiment can usually be designed so that the third postulate is valid though with livestock this cannot always be controlled. The first two postulates seem to be warranted as approximations with respect to many metrical characters governed by many genes each having a relatively small effect; even completely dominant gene differences along with differences showing two factor types of epistatic effects can usually be almost entirely accounted for in terms of additive gene action (Wright, 1935; Lush, 1945). For random mating the correlation between full sibs is

$$\frac{1}{2}\frac{\sigma_G^2}{\sigma_H^2} + \frac{1}{4}\frac{\sigma_D^2}{\sigma_H^2} + r_{I_j I_k}\frac{\sigma_I^2}{\sigma_H^2}$$

where σ_H^2 is the total genetic variance; σ_G^2 the additively genetic variance; σ_D^2 the dominance variance; σ_I^2 the epistatic variance and $r_{I_j I_k}$ the correlation between the epistatic deviations of two siblings (Wright, 1935). Several interesting tables appear in Wright's 1952 paper showing the effects of dominance for varying gene frequencies on variances and correlations as well as one showing an analysis of the variability of two-factor F_2's* in which the 9:3:3:1 ratio is modified in different ways. An experiment having to do with models involving dominance will be discussed somewhat later.

Now under the assumptions stated above, portions of the genetic variance are contained in σ_f^2 and σ_m^2, the variances arising from differences in dams and sires, respectively, obtained in the analysis of variance. In addition because of segregation a further portion of the genetic variance is contained in σ_0^2, the component of variance for individuals within full sib families.

The environmental variance may consist of random effects entirely so that it is all contained in σ_0^2 or it may contain, in addition, differences between litters within the same full sib family in which case we will have a corresponding litter contribution, σ_1^2, and finally it may contain differences between paternal half sibs due to differences in mothering ability of dam, age of dam, etc.—so-called maternal effects, which will be contained in σ_f^2.

If sex linkage, a particular kind of non-allelic interaction, is operating we may have a reduction either in σ_m^2 or in the genetic portion of σ_f^2, depending on which sex is heterogametic and also on the relative effects of gene substitutions in X chromosomes of the two sexes. To take a simple example, if we are analyzing a trait expressed only in females for a population in which the female is the heterogametic sex

---

*Offspring of matings of individuals heterozygous with respect to each of two pairs of genes.

and there are no maternal effects, σ_f^2 will be less than σ_m^2. However, since metric traits are controlled by the relatively small effects of many genes the effect of any sex linkage is likely to be very small in most cases and generally obscured by sampling errors in σ_f^2 and σ_m^2.

Let us consider a population consisting of the mnk progeny of m sires each mated at random to n dams. We can now analyze and interpret the variances as follows, for a trait about which we can make the four assumptions previously stated:

Mean Squares	Expected values	Interpretation
MS_m	$\sigma_0^2 + k\sigma_f^2 + nk\sigma_m^2$	$\sigma_m^2 = r_{hs}^G \sigma_G^2$
MS_f	$\sigma_0^2 + k\sigma_f^2$	$\sigma_f^2 = (r_{fs}^G - r_{hs}^G)\sigma_G^2$
MS_0	σ_0^2	$\sigma_0^2 = (1 - r_{fs}^G)\sigma_G^2 + \sigma_E^2$

Since:

$$1 - r_{fs}^G \frac{\sigma_G^2}{\sigma_T^2} = 1 - \rho_{fs} = \frac{\sigma_0^2}{\sigma_T^2} \; ;$$

$$(r_{fs}^G - r_{hs}^G) \frac{\sigma_G^2}{\sigma_T^2} = \rho_{fs} - \rho_{hs} = \frac{\sigma_f^2}{\sigma_T^2} \; ; \qquad r_{hs}^G \frac{\sigma_G^2}{\sigma_T^2} = \rho_{hs} = \frac{\sigma_m^2}{\sigma_T^2}$$

where ρ_{fs} and ρ_{hs} are the phenotypic intraclass correlations for full sibs and half sibs, respectively, r_{fs}^G and r_{hs}^G are their genetic correlations and

$$\sigma_m^2 + \sigma_f^2 + \sigma_0^2 = \sigma_T^2 = \sigma_G^2 + \sigma_E^2 \; .$$

If we can assume random mating $r_{hs}^G = 1/4$ and $r_{fs}^G = 1/2$; if our population were an inbred line the values would be different. If the values of σ_m^2 and σ_f^2 were significantly different and we were dealing with a trait for which we suspected environmental differences between dam means that is, maternal effects, we would have $\sigma_f^2 = (r_{fs}^G - r_{hs}^G) \sigma_G^2 + \sigma_M^2$. I have kept this example simple so that the relationships would be clear. Numerous papers have been written covering much more complex analyses, some of which are included among the references.

Success in mass selection for improvement of a trait and probably to a large extent for family selection as well, in the absence of intense inbreeding, depends upon the ratio of the additively genetic variance to the total variance, the heritability, usually denoted by h^2. Thus we must have some idea of the importance of non-additive effects. In practice this probably cannot be obtained from the above type of analysis of variance except that in theory advantage could be taken

of the expectation that correlation between sufficiently distant relatives would be due entirely to additive variance whereas nonadditive genetic variance would contribute to the correlation between close relatives. However, we can resort to the comparison of the parent-offspring phenotypic correlation with the full sib phenotypic correlation, for the former will equal 1/2 the ratio of the additively genetic to the total variance while the latter includes, in addition, 1/4 the ratio of the dominance variance to the total variance. Intense selection leads to assortative mating in which case the regression of offspring on mid-parent should be used in place of the correlation; however with dominance present in any degree assortative mating introduces a correlation between the dominance deviations of parents and of offspring and between dominance deviations of either and additive deviations of the other so that the accurate estimation of the degree of heritability becomes practically impossible.

The detection of non-additive gene action from the relations of components of variance for just one generation of an actual population has, as far as I know, been attacked only by Comstock and Robinson in a series of related papers on the estimation of the average degree of dominance for a multigenic trait. (1948, 1949; 1952). They have done extensive work on the appropriate design for estimating a measure of dominance "a" which they define as

$$\frac{(Aa - AA) + (Aa - aa)}{(AA - aa)}$$

while $AA - aa = u$. For two designs the experimental material consists of progeny from random matings among plants of the F_2 generation of a cross of two nearly isogenic lines; in Experiment 1 each of sm male parents is mated with n different female parents while in Experiment 2 all of the mn possible matings of m males and n females in each of s sets are made. In both cases there are s sets of progeny from smn matings in a randomized block arrangement with plot replications. The experimental material in the third design consists of s sets of n pairs of progenies, the members of each pair having the same F_2 male parent but different female parents from each of the two inbred lines which produced F_2. The assumptions made in deriving the genetic interpretations of variance components they state to be fulfilled with two exceptions: (1) no epistasis and (2) no linkage among genes affecting the trait or, if linkages exist, the distribution of genotypes is at equilibrium with respect to coupling and repulsion phases. They point out that the failure of these assumptions to be valid causes an upward bias in their estimates of "a".

They defined the additive genetic variance for the ith locus as that portion of the variance of genetic effects explained by the regression of the genetic effect, y, on the number, x, of A (or a) genes in the genotype, and the dominance variance as the variation of deviations from that regression. With random matings and a frequency of .5 for A for all loci at which there was segregation they derived for n genes in Experiment 1 the following expressions:

$$\sigma_G^2 = \tfrac{1}{2}\Sigma u_i^2 \qquad \sigma_D^2 = \tfrac{1}{4}\Sigma a_i^2 u_i^2$$

$$\left(\frac{2\sigma_D^2}{\sigma_G^2}\right)^{1/2} = \left(\frac{\Sigma a_i^2 u_i^2}{\Sigma u_i^2}\right)^{1/2} = (\bar{a}^2)^{1/2}$$

$\bar{a}^2$ is thus the mean of the a^2's for all loci, weighted relative to the u^2's for the corresponding loci. The variances arising from differences in males, σ_m^2, is shown to be equal to $\sigma_G^2/4$ and that arising from differences in females, σ_f^2, equal to $(\sigma_G^2 + \sigma_D^2)/4$. σ_m^2 and σ_f^2, components for the mean square expectations, contain only genetic variance under these experimental conditions and for a trait having no maternal effects; hence

$$\left(\frac{2(\sigma_f^2 - \sigma_m^2)}{\sigma_m^2}\right)^{1/2} = (\bar{a}^2)^{1/2} \qquad \text{and} \qquad \hat{a} = \left(\frac{2(\hat{\sigma}_f^2 - \hat{\sigma}_m^2)}{\hat{\sigma}_m^2}\right)^{1/2}$$

is an estimate of $\bar{a}$. $\hat{a}$ for experiments 2 and 3 is

$$\left(\frac{2\hat{\sigma}_{mf}^2}{\hat{\sigma}_f^2}\right)^{1/2} \qquad \text{and} \qquad \left(\frac{\hat{\sigma}_{m1}^2}{2\hat{\sigma}_m^2}\right)^{1/2},$$

respectively, where σ_{mf}^2 is the progeny variance due to interaction of male and female parents, σ_f^2 and σ_m^2 are as defined above and σ_{m1}^2 is the progeny variance due to interaction of F_2 and inbred parents. An estimate of σ_G^2, the additive genetic variance, is also obtained from the data of these experiments.

The authors point out that $\hat{a}$ will be somewhat larger than $\bar{a}$, since at least some a's are unequal, and suggest that the bias might be large if some a's were positive and others negative, since it is the average absolute magnitude of a that is being estimated. If $\bar{a}$ is significantly greater than 1, at least one of the a's must be greater than 1 so that overdominance at one or more loci is indicated. If the assumptions of equilibrium with respect to segregation of linked genes and no interallelic interactions do not hold, $\bar{a}$ may be significantly greater than 1 even when there is no overdominance.

Experiment 3 does not depend upon having gene frequencies of .5

and in this case $\bar{a}^2$ is

$$\frac{\Sigma q_i(1 - q_i)a_i^2 u_i^2}{\Sigma q_i(1 - q_i)u_i^2}.$$

Thus the weighting of the a's depends to some extent on shifts in gene frequency which may be variable by loci. The upward bias due to linkage will again be present; however, if the bias declines rapidly as opportunity is provided for recombination, Experiment 3 offers a means of measuring that decline since the probability of an estimate significantly greater than 1 is a function of the expected value of the estimate rather than of $\bar{a}$ when the two are not equal. It is suggested that the apparent overdominance possible in these estimates of $\bar{a}$ has much the same significance for short-run breeding practice as true overdominance.

An exact F test for Experiment 3 and approximate F tests for Experiments 1 and 2 are presented: for example, in testing for overdominance we test whether $\bar{a}$ is significantly greater than 1; if we want to establish the conclusion that the various loci exhibit no dominance or only partial dominance, we test whether $\bar{a}$ is significantly less than 1. The F tests are essentially tests of whether one mean square differs significantly from an estimate of this mean square based on a linear combination of other mean squares. The estimate used is such that its expected value is equal to that of the mean square it is estimating when $\bar{a}$ is equal to 1.

It is shown that, as would be expected, Experiment 2, when the experimental material permits its use, is better than Experiment 1 since the estimate $\hat{\sigma}_D^2$ depends on mean squares with fewer degrees of freedom in the first experiment. Experiment 3 is shown to be the most powerful, the plot requirement being 1/12 to 1/10 as great as for Experiment 1 and 1/4 to 1/2 as great for Experiment 2.

Statistical Techniques

Most of the published papers on estimating variance components are concerned with the one-way classification, nested classifications and with factorial classifications having equal sub-class numbers. Data from breeding experiments often, in fact usually, involve unequal numbers of classes and class numbers. This causes no real trouble when we are dealing with nested classifications until we reach the point of estimating errors but does create difficulties in factorial experiments. Furthermore, we are frequently dealing with the Mixed Model in which some of the effects are assumed to be fixed rather than random variables. Biometrics 1953 presented a paper by Henderson which

has satisfied a real need. Here Henderson discusses three methods for estimating variance components under the above mentioned handicaps and illustrates their application with some genetic data. I shall outline the methods and give his conclusions concerning them.

Method 1 consists in computing sums of squares as in the corresponding orthogonal case, equating these sums of squares to their expectations, derived under the assumptions of the Eisenhart Model II, and solving for the unknown components. Formulas for the computation of the coefficients of the various components of variance in the expected values of the different sums of squares are given. Method 1 is the simplest but gives biased estimates of some of the components when we are dealing with the Mixed Model while assuming Model II. Another bias is present if some of the elements of the Model are correlated. Method 2 is again not difficult. The model Henderson has taken is as follows:

$$Y_{hijk} = \mu + a_h + h_i + s_j + (hs)_{ij} + \epsilon_{hijk}$$

where the a_h's, $h = 1, 2, \cdots p$, are fixed effects. Least squares estimates of the a's and of the d_{ij}'s [$d_{ij} = \mu + h_i + s_j + (hs)_{ij}$] are estimated jointly. The least square equations reduce to p in number and, with the imposition of one restriction on the a's, reduce again to $p - 1$ in $p - 1$ unknowns. Solutions for the $\hat{a}_h$ are obtained as well as the inverse of the $p - 1$ rowed matrix used in the solutions. This inverse, in turn, is used in combination with different matrices formed from the various class or subclass numbers to estimate the corrected coefficients for the variance components corresponding to the corrected sums of squares. The latter are obtained from new class totals, corrected for the a's. The Σy_{hijk}^2 is corrected by the reduction $R(a_h, d_{ij})$ and all components are then estimated. Method 2 gives estimates which are free from the bias resulting from using Method 1 when some of the effects are fixed but is still biased when some of the effects are correlated.

Method 3, which is unbiased but formidable, consists in computing the mean squares by a conventional least squares analysis (method of fitting constants, for example) of nonorthogonal data, equating these mean squares to their expectations and solving for the unknown variances. As in Method 2, the inversions of certain matrices are required in order to obtain the coefficients of the variance components in the expectations. The relative sizes of the sampling variances of estimates obtained by the three methods are not known.

An excellent report of the progress which has been made in the estimation of variance components and of sampling variances of these estimates has been presented by Crump (1951) who also indicates

situations for which the sampling variances are not known. All too frequently people working in animal and plant breeding find themselves facing just such a situation. Since estimates of genetic parameters, for example, genetic variance, dominance variance, heritability, etc., are functions of one or more mean squares from the analysis of variance, estimation of the sampling variances of such estimates is difficult at best.

Crump defines a balanced classification as one in which all of the classes or subclasses of any chosen rank contain the same number of observations. If we consider first a balanced multiple classification for Model II with degrees of freedom which are not very large and we are interested in estimating the sampling variance of an estimate of a genetic parameter,

$$\hat{\sigma}^2 = a_1 M_1 + a_2 M_2 + \cdots$$

where M_i is a mean square with degrees of freedom r_i, several methods of attack are open to us. Satterthwaite (1941, 1946) has examined the distribution of $\hat{\sigma}^2$ and has recommended that it be approximated by a χ^2 distribution with effective degrees of freedom, r, determined by the relation

$$r = \frac{[a_1 M_1 + a_2 M_2 + \cdots]^2}{\dfrac{(a_1 M_1)^2}{r_1} + \dfrac{(a_2 M_2)^2}{r_2} + \cdots}$$

He suggests that the approximation should be used cautiously when one or more of the a's is negative since the approximating distribution does not allow negative values of $\hat{\sigma}^2$. An approximation of this type was suggested by H. F. Smith (1936) for a problem involving only two mean squares with $a_1 = a_2 = 1$. Bross (1950) has constructed an approximate fiducial interval for a variance component, σ_b^2, arising from the class differences in a one-way classification based upon Fisher's solution (1935). The limits are functions of $\hat{\sigma}_b^2$, F obtained from the data and tabular values of F_α. Bross also gives approximate confidence intervals for σ_b^2 by using the fact that when $\sigma_b^2 \neq 0$ (M_b/M_w) is distributed as $F[E(M_b)]/[E(M_w)]$ so that, if $E(M_w) = \sigma_0^2$ and $E(M_b) = \sigma_0^2 + n\sigma_b^2$, $(F/F_\alpha - 1)\,\sigma_0^2/n$ is an exact lower confidence limit for σ_b^2 and $(F/F_\alpha - 1)\,(\hat{\sigma}_b^2/F - 1)$ is a rough lower confidence limit. Both Satterthwaite and Bross investigated to some extent the accuracy of their approximations but, as Crump points out, more investigation is needed.

Cochran (1951) presents an approximate F test,

$$F' = \frac{M_1 + M_2 + \cdots + M_r}{M_{r+1} + M_{r+2} + \cdots + M_k}$$

for a linear relation among variances, $\theta_1 + \theta_2 + \cdots + \theta_r = \theta_{r+1} + \theta_{r+2} + \cdots + \theta_k$ using as one of his illustrative problems that of Comstock and Robinson discussed previously in which the expectations of three mean squares were connected by a linear relation, the coefficients being functions of the quantity $\bar{a}$ which was used as a measure of the degree of dominance. Hence the null hypothesis that $\bar{a}$ is not different from a specified value leads to known values for the coefficients and the linear relation can be tested from the data. The effective degrees of freedom, n_1' and n_2' for Cochran's F' test are found by the rule suggested by Smith and Satterthwaite. Of the possible F' ratios which could be formed, Cochran suggests using one for which the coefficients of the mean squares are positive. He investigates and recommends the F' test in the case of three variances only, $\theta_3 = \theta_1 + \theta_2$, affirming that the approximation would be less satisfactory with four variances since two nuisance parameters of the type θ_i/θ_j would be involved.

If we take up the unbalanced case, which will be the one most likely to be encountered in animal and plant work we find ourselves very much in the dark with respect to the reliability of estimates of genetic parameters. Consider first the one-way classification under Model II: $y_{ij} = \alpha_i + \epsilon_{ij} + \mu$ with the variances of the normally distributed α_i's and ϵ_{ij}'s being σ_α^2 and σ_ϵ^2 respectively. We observe a class of N_i individuals. Now the within class mean square will be distributed like χ^2 but the between class mean square, while independent of the former, will not have an ordinary χ^2 distribution when σ_α^2 is $\neq 0$. In addition, Crump (1951) points out that $\hat{\sigma}_\epsilon^2 = M_w$ and $\hat{\sigma}_\alpha^2 = (M_b - M_w)/N_0$ are not maximum likelihood estimates of σ_ϵ^2 and σ_α^2. N_0 is the coefficient of σ_α^2 in $E(M_b)$ and is equal to $1/(a-1) [\Sigma N_i - \Sigma N_i^2/\Sigma N_i]$. Crump has derived the sampling variances of $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\alpha^2$ as well as those of $\tilde{\sigma}_\epsilon^2$ and $\tilde{\sigma}_\alpha^2$, the maximum likelihood estimates of σ_ϵ^2 and σ_α^2. He shows that $V(\hat{\sigma}_\epsilon^2)$, the variance of $\hat{\sigma}_\epsilon^2$, approaches that of $\tilde{\sigma}_\epsilon^2$, $V(\tilde{\sigma}_\epsilon^2)$, as the numbers in the classes increase, independently of a, and points out that $V(\hat{\sigma}_\alpha^2)$ has a low efficiency relative to $V(\tilde{\sigma}_\alpha^2)$ when a, the number of classes, is small though the ratio $V(\tilde{\sigma}_\alpha^2)/V(\hat{\sigma}_\alpha^2)$, proved to be so complex that he was unable to study its behavior. Tukey (1950) estimates σ_ϵ^2 by M_w, and σ_α^2 by

$$\frac{1}{a-1}\left\{\Sigma\bar{y}_i^2 - \frac{(\Sigma\bar{y}_i)^2}{a}\right\} - \left\{\frac{1}{a}\Sigma\frac{1}{N_i}\right\}M_w$$

and derives the sampling variances which, according to Crump have not yet been compared with $V(\hat{\sigma}_\alpha^2)$ and $V(\tilde{\sigma}_\alpha^2)$.

Apparently no sampling variances have been derived in the unbalanced case for multiple classifications under Model II or the Mixed

Model and these are of course just what are needed if we are to estimate sampling variance for the estimate of the genetic variance, $\hat{\sigma}_G^2$, obtained from an unbalanced classification corresponding to the scheme presented previously.   An estimate $c_1 \hat{\sigma}_m^2 + c_2\hat{\sigma}_f^2$ of σ_G^2 can be derived from that scheme but until we know something of the sampling variances we lack criteria for choosing those values of c_1 , and c_2 which will give the best estimate of h^2.   Osborne (1952) has published approximate sampling variances for several estimates of h^2 but as these are functions of the unknown sampling variances of the components $\hat{\sigma}_m^2$, $\hat{\sigma}_f^2$ and $\hat{\sigma}_0^2$ his approximations could be poor in the commonly occurring unbalanced case.   Comstock and Robinson in their experimental work on the consistency of estimates of variance components in a balanced design, point out that their results cannot speak for other sorts of data.

Wald (1940, 1941) gives a method for placing confidence limits on the ratio of any variance component to the error component for the unbalanced case in multiple classifications under Model II.   For the one-way classification, for example, Wald showed that:

$$F = \frac{N - a}{a - 1} \frac{\Sigma w_i\left[\bar{y}_i - \dfrac{\Sigma w_i\bar{y}_i}{\Sigma w_i} \right]^2}{\Sigma\Sigma(y_{ij} - \bar{y}_i)^2} \text{ , where } w_i = \frac{N_i}{1 + N_i\lambda^2} \text{ and } \lambda^2 = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \text{ ,}$$

has the analysis of variance F distribution with $a - 1$ and $N - a$ degrees of freedom.   Thus the lower confidence limit is given by the root of the following equation in λ^2

$$\frac{N - a}{a - 1} \frac{\Sigma w_i\left[\bar{y}_i - \dfrac{\Sigma w_i\bar{y}_i}{\Sigma w_i} \right]^2}{\Sigma\Sigma(y_{ij} - \bar{y}_i)^2} = F_{u_a}$$

Wald shows that each of the two equations, one for F_{u_a} and one for F_{1_a} have at most one root in λ^2 ; that if one equation has no root the corresponding confidence limit must be set equal to 0 and that if neither has a root we must reject one of the hypotheses:

$y_{ij} = \alpha_i + \epsilon_{ij} + \mu$
ϵ_{ij} and α_i are normally and independently distributed
Each ϵ_{ij} has the same distribution
Each α_i has the same distribution.

Solutions of such equations are obviously difficult in practice but if they are obtained for the particular types of lack of balance with which one is accustomed to work, one would have some idea of the accuracy of approximations he may be using.

I should like to add in closing that an unbalanced classification in

which the subclass numbers are correlated genetically with the variate studied will give rise to further difficulties and that this is not an unlikely situation for certain traits.

LITERATURE CITED AND REFERENCES

Bross, Irwin, 1950. Fiducial intervals for variance components. *Biometrics 6*: 136–144.

Byswaters, J. H., 1937. The hereditary and environmental portions of the variance in weaning weights of Poland China Pigs. *Genetics 22*: 457–468.

Cochran, W. G., 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biom. Bull. 3*: 22–38.

——————, 1951. Testing a linear relation among variances. *Biometrics 7*: 17–32.

Comstock, R. E. and H. F. Robinson, 1948. The components of genetic variance in populations of bi-parental progenies and their use in estimating the average degree of dominance. *Biometrics 4*: 254–266.

——————, ——————, and P. H. Harvey, 1949. A breeding procedure designed to make maximum use of both general and specific combining ability. *Agron. Jour. 41*: 360–367.

——————, and ——————, 1952. Estimation of average dominance of genes. *Heterosis*, edited by John W. Gowen, Iowa State College Press, Ames, Iowa.

Crump, S. L., 1946. The estimation of variance components in analysis of variance. *Biom. Bull. 2*: 7–11.

——————, 1951. The present status of variance component analysis. *Biometrics 7*: 1–16.

Daniels, H. E., 1939. The estimation of components of variance. *J. Roy. Stat. Soc. Supp. 6*: 186–197.

Eisenhart, Churchill, 1947. The assumptions underlying analysis of variance. *Biom. Bull. 3*: 1–21.

Fieller, E. C. and C. A. B. Smith, 1951. Note on the analysis of variance and intraclass correlation. *Ann. Eugen. 16 Part 1*: 97–104.

Fisher, R. A., 1918. The correlations between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh 52*: 399–433.

——————, 1925. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh and London.

——————, 1935. The fiducial argument in statistical inference. *Ann. Eugen. 6*: 391–398.

Hetzer, H. O., G. E. Dickerson and J. H. Zeller, 1944. Heritability of type in Poland China swine evaluated by scoring. *Jour. An. Sci. 3*: 390–398.

Henderson, C. R., 1953. Estimation of variance and covariance components. *Biometrics 9*: 226–252.

Knapp, Bradford Jr. and A. W. Nordskog. 1946. Heritability of growth and efficiency of beef cattle. *J. An. Sci. 5*: 62–70.

Lerner, I. M., 1950. *Population genetics and animal improvement*. University Press, Cambridge.

Lush, J. L., 1945. *Animal breeding plans*. The Collegiate Press. Ames, Iowa.

Mather, K., 1949. *Biometrical genetics*. Dover Publications Inc.

Osborne, R. and W. S. B. Paterson, 1952. On the sampling variance of heritability estimates derived from variance analysis. *Proc. Roy. Soc. Edinburgh, B, LXIV*: 456–461.

Satterthwaite, F. E., 1941. Synthesis of variance. *Psychometrica VI*: 309–316.
————, 1946. An approximate distribution of estimates of variance components. *Biom. Bull. 2*: 110–114.
Smith, H. F., 1936. The problem of comparing the results of two experiments with unequal errors. *Jour. G. S. I. R. (Australia) 9*: 211–212.
Stonaker, H. H. and J. L. Lush, 1942. Heritability of conformation in Poland China swine as evaluated by scoring. *J. An. Sci. 1*: 99–105.
Tukey, J. W., 1950. Finite sampling simplified. *Memorandum report 45 of the Statistical Research Group*, Princeton University.
Wald, A., 1940. A note on the analysis of variance with unequal class frequencies. *Ann. Math. Stat. 11*: 96–100.
————, 1941. On the analysis of variance in case of multiple classifications with unequal class frequencies. *Ann. Math. Stat. 12*: 346–349.
Weinberg, W., 1910. Weitere Beiträge zur Theorie der Vererbung. *Arch. Rass.-u Ges. Biol. 7*: 35–49, 169–173.
Wright, S., 1917. The average correlation within sub-groups of a population. *J. Wash. Acad. Sci. 7*: 532–535.
————, 1918. On the nature of size factors. *Genetics 3*: 367–374.
————, 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proc. Nat. Acad. Sci. 6*: 320–332.
————, 1921. Correlation and causation. *J. Agric. Res. 20*: 557–585. Systems of mating. *Genetics 6*.
————, 1935. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *J. Genetics 30*: 243–256.
————, 1952. *The genetics of quantitative variability.* Agricultural Research Council, London: Her Majesty's Stationery Office.

FITTING THE NEYMAN TYPE A (TWO PARAMETER) CONTAGIOUS DISTRIBUTION

J. B. Douglas

School of Mathematics,
N.S.W. University of Technology, Australia.

Introduction

One of the difficulties associated with the use of the Neyman contagious distributions (Neyman (1939)) concerns the method of fitting to data. Restricting attention to the two parameter Type A distribution, the method suggested by Neyman (with a remark that its efficiency needed investigation: there are no sufficient estimators) was to equate the corresponding first and second moments of the data and the distribution—this gives two readily soluble equations for the two estimators. Shenton (1949) investigated the efficiency of this moment fit, and outlined a technique for an iterative maximum likelihood fitting process, together with suggestions relating to the circumstances in which the process might be worth applying. Owing to the complicated nature of the recurrence relation for successive probabilities, the distribution is rather tedious to handle in any circumstances, and it is unfortunately the case that the maximum likelihood process suggested by Shenton increases considerably the labour of fitting. Recent papers (e.g., Beall and Rescia (1953)) stress the need for a technique which would reduce the amount of calculation—this paper suggests a method which greatly shortens the labour of obtaining a maximum likelihood fit, and which reduces the calculation necessary for a comparison of observation and expectation whatever the method of fitting used. As with the Shenton technique, the successive approximations for the maximum likelihood fit are based on the Newton-Raphson method.

The case where the zero class is unknown is also briefly discussed.

149

I The Complete Two Parameter Neyman Type A Distribution

The probability of the occurrence of x individuals is given by

$$(1) \qquad P_x = \frac{e^{-\mu}\nu^x}{x!} \sum_{r=0}^{\infty} \frac{(\mu e^{-\nu})^r r^x}{r!} , \quad x = 0, 1, 2, \cdots , \quad 0^x = \begin{cases} 1, x = 0 \\ 0, x > 0, \end{cases}$$

where μ and ν are positive parameters (in Neyman's 1939 paper written m_1 and m_2 respectively, but the suffixes are rather inconvenient when generalizations are not being considered). Successive probabilities are found from the recurrence relation

$$(2) \qquad P_{x+1} = \frac{\mu\nu e^{-\nu}}{x+1} \sum_{s=0}^{x} \frac{\nu^s}{s!} P_{x-s} .$$

For a sample with observed frequencies of f_0 in the zero class, f_1 in the class with one member, $\cdots$, f_x in the class with x members, $\cdots$, and power sums

$$S_r = \sum_x x^r f_x , \text{ say,}$$

the summation being over all observed classes, the moment estimators $\hat{\mu}_m$, $\hat{\nu}_m$ of μ, ν respectively are given by

$$(3) \qquad \hat{\nu}_m = \frac{S_2}{S_1} - \frac{S_1}{S_0} - 1, \qquad \hat{\mu}_m = \frac{S_1}{S_0} \cdot \frac{1}{\hat{\nu}_m} ,$$

in a form convenient for desk machine computation*.

The maximum likelihood estimators $\hat{\mu}$, $\hat{\nu}$ are the solutions of the likelihood equations

$$(4) \qquad \begin{aligned} \mu\nu &= S_1/S_0 = \bar{x}, \quad \text{say,} \\ \sum f_x \pi_x &= S_1 , \quad \text{with} \quad \pi_x = (x+1)P_{x+1}/P_x , \end{aligned}$$

and effectively the procedure suggested by Shenton (1949) was to write

$$(5) \qquad \begin{aligned} F(\nu) &= \sum f_x \pi_x - S_1 , \\ F'(\nu) &= \frac{1}{\nu} \sum f_x \pi_x - \frac{\nu+1}{\nu^2} \sum f_x \chi_x , \quad \chi_x = \pi_x(\pi_{x+1} - \pi_x), \end{aligned}$$

where μ was supposed eliminated through the first likelihood equation, and, with first approximations μ_1 , ν_1 obtained from the moment equa-

---

*Neyman (1939) apparently used an unbiassed second moment estimator. The difference is unimportant for the large sample sizes here dealt with.

tions, to determine second approximations μ_2, ν_2 from

(6)
$$\nu_2 = \nu_1 - F(\nu_1)/F'(\nu_1),$$

$$\mu_2 = \bar{x}/\nu_2.$$

(Shenton's equation (9) has a misprint in it: P_{x+1} in the final term should be P_{x+1}^2.)

Writing the expressions in the above form suggests what is in fact a fairly convenient way of proceeding—tabulating x, f_x, P_x, π_x and $f_x \pi_x$ in turn enables one to calculate all the terms necessary and to maintain simultaneous control of the accuracy, with a reasonably complete check of the calculations. It might be pointed out that control of the accuracy needs some care, owing to the occurrence of ratios and subsequent differencing—it is necessary to carry many significant figures in the early stages (say, in the P_x) in order to retain digits with meaning in the final stages for large values of x (say, in the χ_x); it is very easy to give entirely meaningless digits in the sums for both $F(\nu)$ and $F'(\nu)$.

However, the process requires iteration, very often, and the labour involved is so considerable that this is not likely to be carried out for routine fitting. But it is possible to rewrite some of the preceding work, so that with the provision of suitable Tables the labour can be much reduced—the various P_x need not be calculated, for example, until a direct comparison with the observed frequencies is required, and then only to the accuracy necessary for such a comparison in place of the extreme accuracy required as described above.

For

$$\lambda = \mu e^{-\nu},$$

$$P_0 = e^{\lambda - \mu},$$

$$P_x = e^{-\mu} \frac{\nu^x}{x!} \sum_{r=0}^{\infty} \frac{r^x}{r!} \lambda^r, \qquad x > 0,$$

and writing

(7)
$$\mu_x' = e^{-\lambda} \sum_{r=0}^{\infty} r^x \frac{\lambda^r}{r!},$$

so that μ_x' is the x-th power moment about the origin of a Poisson distribution with parameter λ,

(8)
$$P_x = P_0 \frac{\nu^x}{x!} \mu_x'.$$

Given a table of μ'_x , this means that P_x can be calculated without knowledge of P_{x-1} , P_{x-2} , $\cdots$; in practice, it turns out to be more convenient to tabulate

(9) $$\frac{\mu'_{x+1}}{\mu'_x} = p_x , \qquad \text{say,}$$

whence

(10) $$P_{x+1} = \nu \frac{P_x}{x+1} p_x ,$$

a recurrence relation involving only the immediately preceding probability.

The maximum likelihood equations are then

(11) $$\hat{\mu}\hat{\nu} = \bar{x},$$
$$\hat{\mu} = \frac{1}{S_0} \sum f_x \hat{p}_x ;$$

writing as before

(12) $$F(\nu) = \nu \sum f_x p_x - S_1 ,$$
$$F'(\nu) = \sum f_x p_x - (1 + \nu) \sum f_x q_x \quad \text{with} \quad q_x = p_x(p_{x+1} - p_x),$$

and exactly the procedure outlined above applies. However, given the Table of p_x with interlinear values of q_x , all that need be done is to enter the table with the value of

$$\lambda_1 = \mu_1 e^{-\nu_1}$$

from a moment fit and cumulate $f_x p_x$ and $f_x q_x$, revised estimates of ν and μ being obtained from

$$\nu_2 = \nu_1 - F(\nu_1)/F'(\nu_1) \qquad \text{and} \qquad \bar{x}/\nu_2 ,$$

respectively. Iteration, until no change is produced in the estimates, requires only minutes, compared with hours for a single iteration for the previous process.

To illustrate the procedure, the European Corn Borer data quoted by Neyman (1939) will be used, although in fact one would expect the moment fit to be of reasonably high efficiency (from Shenton's results). Here the moment estimates for μ, ν are 2.21, 1.43, giving an estimate for $\lambda = \mu e^{-\nu}$ of 0.53. Writing the observed frequencies f_x across a slip of paper in the positions corresponding to $x = 0, 1, 2, \cdots 12$ in the Table, cumulation of $f_x p_x$ and $f_x q_x$ leads to revised estimates for μ, ν of 1.98, 1.60. These give a revised estimate for λ of 0.40, which leads

to a further pair of estimates for μ, ν of 2.06, 1.54; use of these (with λ estimated by 0.44) leads again to the same pair of estimates. (No greater accuracy can be obtained from the Table.)

The calculation of the P_x is made from

$$P_0 = e^{\lambda - \mu},$$

$$P_x = \nu \frac{P_{x-1}}{x} p_{x-1}, \qquad x > 0,$$

and the observed frequencies f_x are shown below with the expected frequencies $\phi_x = S_0 P_x$. These may also be compared with the expected frequencies F_x from the moment fit, and with the expected frequencies Φ_x from a maximum likelihood fit with four iterations along the lines set out by Shenton (the final estimates of μ, ν are 2.063, 1.535).

TABLE I

x	0	1	2	3	4	5	6	7	8+	x^2
f_x	24	16	16	18	15	9	6	5	11	—
ϕ_x	23.8	16.1	17.8	16.0	13.1	10.1	7.3	5.1	10.7	1.07
F_x	22.3	16.8	18.4	16.5	13.4	10.3	7.5	5.2	9.6	1.48
Φ_x	23.8	16.2	18.0	16.1	13.2	10.2	7.5	5.3	9.7	1.33

(In each case, the x^2 has 6 degrees of freedom.) The discrepancies between the ϕ_x and Φ_x are small—since in fact the earlier classes are the more important (cf., e.g., Anscombe (1949), (1950)), the tendency for accumulation of errors for large values of x is probably unimportant.

II The Truncated Two Parameter Neyman Type A Distribution

In some circumstances, the zero class is either unreliable or entirely unknown, as for instance when one cannot be sure that all individuals *not* possessing some characteristic have been identified, or where the number of animals *not* trapped even once is quite unknown. (The circumstance of misclassification is not considered—i.e., f_0 is suspect or unknown, but not f_1, f_2, $\cdots$.)

The distribution then appropriate has probabilities P'_x related to the previous P_x by

(13) $$P'_x = \frac{P_x}{1 - P_0}, \qquad x = 1, 2, 3, \cdots,$$

and the moment estimators $\hat{\mu}_m$, $\hat{\nu}_m$ are given by*

(14)
$$\frac{\hat{\mu}_m \hat{\nu}_m}{1 - \hat{P}_{0(m)}} = \bar{x}',$$

$$\hat{\nu}_m (1 + \hat{\mu}_m) = \frac{S_2'}{S_1'} - 1,$$

where

$$\hat{P}_{0(m)} = \exp\left[-\hat{\mu}_m (1 - e^{-\hat{\nu}_m})\right]$$

and

$$S_r' = \sum_{x>0} x^r f_x = \sum{}' x^r f_x , \qquad \text{say,}$$

$$\bar{x}' = S_1'/S_0' .$$

However, explicit solutions for $\hat{\mu}_m$ and $\hat{\nu}_m$ cannot be obtained, nor in fact do positive solutions (required by the physical problem) always exist.

It is possible to use the "analogues" of the moment equations (3) in I:

(15)
$$\hat{\nu}_m' = \frac{S_2'}{S_1'} - \frac{S_1'}{S_0'} - 1, \qquad \hat{\mu}_m' = \frac{S_1'}{S_0'} \cdot \frac{1}{\hat{\nu}_m'} .$$

Because these quite commonly lead to negative estimates, their application is not considerable, except as giving some idea of reasonable first approximations for the processes described below, since it is in fact so easy to calculate their values.

Following the Shenton procedure, a maximum likelihood fit for μ and ν leads in the present case to the pair of equations

(16)
$$\frac{\hat{\mu}\hat{\nu}}{1 - \hat{P}_0} = \bar{x}',$$

$$1 - \hat{P}_0 e^{-\hat{\nu}} = \frac{1}{S_1'} \sum{}' f_x \hat{\pi}_x ,$$

and writing

(17a)
$$F(\nu) = 1 - P_0 e^{-\nu} - \frac{1}{S_1'} \sum{}' f_x \pi_x ,$$

---

*Obtained by J. H. Bennett (1950)—unpublished.

where μ is regarded as being eliminated (though this cannot be done explicitly) by using the former of this pair of equations, (16),

$$(17b) \quad F'(\nu) = e^{-\nu}P_0\left(1 + \mu e^{-\nu} - (1 - e^{-\nu})\frac{\mu}{\nu}\frac{1 - \mu\nu e^{-\nu}P_0'}{1 - \mu(1 - e^{-\nu})P_0'}\right)$$

$$- \frac{1}{\nu S_1'}\sum' f_x\pi_x$$

$$+ \frac{1}{\nu S_1'}\left(1 + \frac{1}{\nu}\frac{1 - \mu\nu e^{-\nu}P_0'}{1 - \mu(1 - e^{-\nu})P_0'}\right)\sum' f_x\chi_x ,$$

where again

$$\pi_x = (x + 1)\frac{P_{x+1}'}{P_x'} = (x + 1)\frac{P_{x+1}}{P_x} ,$$

$$\chi_x = \pi_x(\pi_{x+1} - \pi_x),$$

and

$$P_0' = P_0/(1 - P_0).$$

As before, given first approximations μ_1, ν_1, second approximations μ_2, ν_2 can be found from

$$\nu_2 = \nu_1 - F(\nu_1)/F'(\nu_1),$$

$$(18)$$

$$\mu_2 = \frac{\bar{x}'}{\nu_2}(1 - \exp[-\mu_1(1 - e^{-\nu_2})])$$

$$\left(\text{or} \quad \mu_2 = \frac{\bar{x}'}{\nu_2}(1 - \exp[-\mu_1(1 - e^{-\nu_2})])\right).$$

Once more the labour is considerable, and the use of the device appropriate before leads to the equivalent maximum likelihood equations

$$(19)$$

$$\frac{\hat{\mu}\hat{\nu}}{1 - \hat{P}_0} = \bar{x}',$$

$$\frac{1 - \hat{P}_0 e^{-\hat{\nu}}}{\hat{\nu}} = \frac{\sum' f_x\hat{p}_x}{S_1'} ,$$

with

$$(20)$$

$$F(\nu) = 1 - P_0 e^{-\nu} - \frac{\nu}{S_1'}\sum' f_x p_x ,$$

$$F'(\nu) = e^{-\nu}P_0\left(1 + \mu e^{-\nu} - (1 - e^{-\nu})\frac{\mu}{\nu}\frac{1 - \mu\nu e^{-\nu}P_0'}{1 - \mu(1 - e^{-\nu})P_0'}\right)$$

$$- \frac{1}{S_1'}\sum' f_x p_x$$

$$+ \frac{1}{S_1'}\left(\nu + \frac{1 - \mu\nu e^{-\nu}P_0'}{1 - \mu(1 - e^{-\nu})P_0'}\right)\sum' f_x q_x$$

giving revised estimates as above. The calculations for $\sum' f_x p_x$ and $\sum' f_x q_x$ are again brief, though the substitution in the expression for $F'(\nu)$ is worth arranging systematically and in any case takes some time*.

If it is possible to make a reasonable assumption regarding the frequency in the zero class, a convenient procedure seems to be to use this "completed" sample to obtain moment estimators of μ and ν (as in I) and then to adjust the estimate of μ to satisfy at least roughly the maximum likelihood equation

$$\frac{\mu\nu}{1 - P_0} = \bar{x}'.$$

(Given μ_1, ν_1, a second approximation to μ can be found from

$$\mu_1 - \frac{f(\mu_1)}{f'(\mu_1)},$$

where

$$f(\mu) = 1 - \mu\frac{\nu_1}{\bar{x}'} - \exp[-\mu(1 - e^{-\nu_1})],$$

$$f'(\mu) = -\frac{\nu_1}{\bar{x}'} + (1 - e^{-\nu_1})\exp[-\mu(1 - e^{-\nu_1})].$$

When it is not possible to proceed as above, either of the two earlier methods may be used to obtain first approximations, though it seems to be worth making some effort to try to satisfy at least roughly the maximum likelihood equation of this paragraph.

Whatever the method, the expected frequencies ϕ'_x are found from

$$\phi'_x = S'_0 P'_x, \qquad x = 1, 2, \cdots,$$

and if using the Table, from

$$\phi'_x = \frac{\nu S'_0}{1 - P_0} \cdot \frac{\phi'_{x-1}}{x} p_{x-1}, \qquad x = 1, 2, \cdots$$

$$\left(\text{where} \qquad \phi'_0 = S'_0 P'_0 = S'_0 \frac{e^{\lambda-\mu}}{1 - e^{\lambda-\mu}}\right).$$

As an illustration, unpublished data of leaf counts of Leucopogon Virgatus (supplied by Dr. D. W. Goodall, of the University of Mel-

---

*Rather than to make repeated use of

$$\nu_{r+1} = \nu_r - F(\nu_r)/F'(\nu_r),$$

it may be preferable to retain a constant value for $F'(\nu)$ once a reasonable estimate of ν has been obtained.

bourne) will serve.   There are *a priori* reasons for suspecting that the
zero class is greatly inflated (but not at the expense of neighboring
classes); the analysis above may therefore be applied.   There are no
(positive) moment estimators, so that the analogues (15) of the moment
equations were used and then adjusted as suggested to give first approxi-
mations for μ, ν of 0.293, 2.17.   Use of equations (19) and (20) led to
second approximations 1.10, 1.44 and then to third approximations
1.49, 1.21 (for which $F(\nu) = -0.002$, so that the limit of accuracy of
the Table has practically been reached; the example is rather an awkward
one, in that $F'(\nu)$ is also rather small near the root of $F(\nu)$).   These in
turn give the expected frequencies ϕ'_x shown:

TABLE II.

x	0	1	2	3	4	5	6	7+
f_x	(798)	70	41	33	29	11	7	11
ϕ_x'	(109)	58.6	51.2	36.1	23.2	14.0	8.1	10.8

and this leads to $\chi^2 = 6.8$, with 4 degrees of freedom—a reasonable fit.
(It is interesting, though perhaps not unexpected, that while estimation
with inclusion of the zero class reproduces f_x for $x = 0$ rather well, the
divergence for $x > 0$ is considerable—above, as was anticipated *a priori*
the divergence is largely at $x = 0$.)

III Tables of p_x and q_x

These give ratios of Poisson power moments about the origin, up
to order 20.   More precisely, for

$$\mu'_x = e^{-\lambda} \sum_{r=0}^{\infty} r^x \frac{\lambda^r}{r!} \, ,$$

then

$$p_x = \frac{\mu'_{x+1}}{\mu'_x} \, ,$$

and p_x is tabulated for

$$x = 0(1)19 \quad \text{and} \quad \lambda = 0.000(0.001)0.03(0.01)0.3(0.1)3.$$

Corresponding values of $q_x = p_x(p_{x+1} - p_x)$ are also shown, in each
case with rather more decimals for small values of x, for various reasons.
(E.g., the number of significant digits thus alters less; f_x for small

values of x is relatively large, in many cases, so that more digits are needed when the sum $\sum f_x p_x$ is required to a fixed number of decimals; and linear interpolation is adequate to more decimals for small x, this being the chief factor in deciding at what points to truncate the tabulated values.)

The horizontal lines across the Tables show the regions above which linear interpolation (with respect to λ) for p_x or q_x is hardly adequate to the accuracy of the Tables. (Some "something" of these lines has been carried out—they are approximate rather than exact.) Where such regions are appropriate, or where greater accuracy than the Tables afford may be required, the Tables can still be used to obtain corrected first approximations for use in the direct Shenton procedure (equations (4), (5) and (6); or (16), (17) and (18)).

References

Anscombe, F. J. (1949). *Biometrics 5*, p. 165.
Anscombe, F. J. (1950). *Biometrika 37*, p. 358.
Beall, G. and Rescia, R. R. (1953). *Biometrics 9*, p. 354.
Neyman, J. (1939). *Annals Math. Stat. 10*, p. 35.
Shenton, L. R. (1949). *Biometrika 36*, p. 450.

APPENDIX—TABLE I

λ	p_0	q_0	p_1	q_1	p_2	q_2	p_3	q_3
0.0	0.000	0.000	1.000	0.000	1.000	0.000	1.00	0.000
0.1	0.100	0.100	1.100	0.100	1.191	0.183	1.34	0.300
0.2	0.200	0.200	1.200	0.200	1.367	0.339	1.61	0.491
0.3	0.300	0.300	1.300	0.300	1.531	0.477	1.84	0.639
0.4	0.400	0.400	1.400	0.400	1.686	0.604	2.04	0.765−
0.5	0.500	0.500	1.500	0.500	1.833	0.722	2.23	0.880
0.6	0.600	0.600	1 600	0.600	1.975+	0.834	2.40	0.989
0.7	0.700	0.700	1.700	0.700	2.112	0.942	2.56	1.095+
0.8	0.800	0.800	1.800	0.800	2.244	1.047	2.71	1.198
0.9	0.900	0.900	1.900	0.900	2.374	1.149	2.86	1.299
1.0	1.000	1.000	2.000	1.000	2.500	1.250+	3.00	1.400
1.1	1.100	1.100	2.100	1.100	2.624	1.349	3.14	1.500
1.2	1.200	1.200	2.200	1.200	2.745+	1.448	3.27	1.599
1.3	1.300	1.300	2.300	1.300	2.865+	1.546	3.40	1.697
1.4	1.400	1.400	2.400	1.400	2.983	1.643	3.53	1.795+
1.5	1.500	1.500	2.500	1.500	3.100	1.740	3.661	1.893
1.6	1.600	1.600	2.600	1.600	3.215+	1.837	3.787	1.991
1.7	1.700	1.700	2.700	1.700	3.330	1.933	3.910	2.088
1.8	1.800	1.800	2.800	1.800	3.443	2.030	4.032	2.186
1.9	1.900	1.900	2.900	1.900	3.555+	2.126	4.153	2.283
2.0	2.000	2.000	3.000	2.000	3.667	2.222	4.273	2.380
2.1	2.100	2.100	3.100	2.100	3.777	2.319	4.391	2.477
2.2	2.200	2.200	3.200	2.200	3.888	2.415−	4.509	2.574
2.3	2.300	2.300	3.300	2.300	3.997	2.511	4.625+	2.671
2.4	2.400	2.400	3.400	2.400	4.106	2.608	4.741	2.768
2.5	2.500	2.500	3.500	2.500	4.214	2.704	4.856	2.865−
2.6	2.600	2.600	3.600	2.600	4.322	2.801	4.970	2.962
2.7	2.700	2.700	3.700	2.700	4.430	2.897	5.084	3.058
2.8	2.800	2.800	3.800	2.800	4.537	2.994	5.197	3.155−
2.9	2.900	2.900	3.900	2.900	4.644	3.098	5.309	3.252
3.0	3.000	3.000	4.000	3.000	4.750	3.188	5.421	3.349

APPENDIX—TABLE I (*Continued*)

λ	p_4	q_4	p_5	q_5	p_6	q_6	p_7	q_7
0.0	1.00	0.000	1.00	0.000	1.00	0.00	1.00	0.00
0.1	1.57	0.421	1.84	0.498	2.11	0.53	2.36	0.55+
0.2	1.92	0.601	2.23	0.658	2.53	0.70	2.80	0.75+
0.3	2.19	0.737	2.53	0.797	2.84	0.86	3.14	0.92
0.4	2.42	0.859	2.77	0.928	3.11	1.00	3.43	1.07
0.5	2.62	0.975+	2.99	1.053	3.35−	1.13	3.69	1.21
0.6	2.81	1.087	3.20	1.174	3.56	1.26	3.92	1.35−
0.7	2.99	1.196	3.39	1.291	3.77	1.38	4.14	1.47
0.8	3.15+	1.305−	3.57	1.405−	3.96	1.50+	4.34	1.59
0.9	3.31	1.411	3.74	1.516	4.14	1.62	4.53	1.71
1.0	3.47	1.515+	3.90	1.626	4.32	1.73	4.72	1.83
1.1	3.62	1.619	4.06	1.733	4.49	1.84	4.90	1.94
1.2	3.76	1.722	4.22	1.840	4.66	1.95+	5.07	2.05+
1.3	3.90	1.824	4.37	1.944	4.82	2.06	5.24	2.17
1.4	4.04	1.926	4.52	2.049	4.97	2.16	5.41	2.27
1.5	4.178	2.027	4.663	2.152	5.12	2.27	5.57	2.38
1.6	4.312	2.127	4.806	2.255−	5.27	2.38	5.73	2.49
1.7	4.444	2.227	4.945+	2.357	5.42	2.48	5.88	2.60
1.8	4.574	2.327	5.083	2.458	5.57	2.58	6.03	2.70
1.9	4.703	2.425+	5.219	2.559	5.71	2.69	6.18	2.81
2.0	4.830	2.524	5.352	2.660	5.85−	2.79	6.33	2.91
2.1	4.955+	2.623	5.485−	2.760	5.99	2.89	6.47	3.01
2.2	5.080	2.721	5.615+	2.860	6.12	2.99	6.61	3.12
2.3	5.203	2.819	5.745−	2.959	6.26	3.09	6.75+	3.22
2.4	5.325−	2.917	5.873	3.059	6.39	3.19	6.89	3.32
2.5	5.446	3.015+	6.000	3.158	6.53	3.29	7.03	3.42
2.6	5.566	3.113	6.125+	3.257	6.66	3.39	7.17	3.52
2.7	5.685+	3.211	6.250	3.356	6.79	3.49	7.30	3.63
2.8	5.804	3.309	6.374	3.454	6.92	3.59	7.44	3.73
2.9	5.922	3.406	6.497	3.553	7.04	3.69	7.57	3.83
3.0	6.039	3.504	6.619	3.650+	7.17	3.79	7.70	3.93

APPENDIX—TABLE I (*Continued*)

λ	p_8	q_8	p_9	q_9	p_{10}	q_{10}	p_{11}	q_{11}
0.0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.1	2.59	0.59	2.82	0.64	3.05−	0.69	3.27	0.73
0.2	3.07	0.81	3.34	0.87	3.60	0.92	3.85+	0.97
0.3	3.44	0.99	3.72	1.05−	4.01	1.11	4.28	1.16
0.4	3.74	1.14	4.05−	1.21	4.35−	1.27	4.64	1.33
0.5	4.01	1.29	4.34	1.36	4.65−	1.43	4.96	1.49
0.6	4.26	1.42	4.60	1.50+	4.92	1.57	5.24	1.64
0.7	4.49	1.55+	4.84	1.63	5.17	1.71	5.51	1.78
0.8	4.71	1.68	5.06	1.76	5.41	1.84	5.75+	1.92
0.9	4.91	1.80	5.28	1.89	5.64	1.97	5.99	2.05+
1.0	5.11	1.92	5.48	2.01	5.85+	2.10	6.21	2.18
1.1	5.30	2.04	5.68	2.13	6.06	2.22	6.42	2.31
1.2	5.48	2.15+	5.87	2.25−	6.26	2.34	6.63	2.43
1.3	5.66	2.27	6.06	2.36	6.45−	2.46	6.83	2.55−
1.4	5.83	2.38	6.24	2.48	6.63	2.58	7.02	2.67
1.5	6.00	2.49	6.41	2.59	6.82	2.69	7.21	2.79
1.6	6.16	2.60	6.58	2.70	6.99	2.80	7.39	2.90
1.7	6.32	2.71	6.75−	2.81	7.17	2.92	7.57	3.01
1.8	6.48	2.81	6.91	2.92	7.34	3.03	7.75−	3.13
1.9	6.63	2.92	7.07	3.03	7.50+	3.14	7.92	3.24
2.0	6.79	3.03	7.23	3.14	7.67	3.25−	8.09	3.35+
2.1	6.94	3.13	7.39	3.25−	7.83	3.35+	8.26	3.46
2.2	7.08	3.24	7.54	3.35+	7.99	3.46	8.42	3.57
2.3	7.23	3.34	7.69	3.46	8.14	3.57	8.58	3.68
2.4	7.37	3.44	7.84	3.56	8.30	3.68	8.74	3.79
2.5	7.52	3.55−	7.99	3.67	8.45−	3.78	8.90	3.89
2.6	7.66	3.65−	8.14	3.77	8.60	3.89	9.05+	4.00
2.7	7.80	3.75+	8.28	3.87	8.75−	3.99	9.20	4.11
2.8	7.94	3.85+	8.42	3.98	8.89	4.10	9.35+	4.21
2.9	8.07	3.96	8.56	4.08	9.04	4.20	9.50+	4.32
3.0	8.21	4.06	8.70	4.18	9.18	4.30	9.65+	4.42

APPENDIX—TABLE I (*Continued*)

λ	p_{12}	q_{12}	p_{13}	q_{13}	p_{14}	q_{14}	p_{15}	q_{15}
0.0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.1	3.49	0.76	3.71	0.80	3.93	0.84	4.14	0.87
0.2	4.10	1.01	4.35+	1.06	4.59	1.10	4.83	1.15−
0.3	4.55+	1.22	4.82	1.27	5.08	1.32	5.34	1.37
0.4	4.93	1.39	5.21	1.45+	5.49	1.51	5.76	1.56
0.5	5.26	1.56	5.55+	1.62	5.85−	1.68	6.13	1.74
0.6	5.55+	1.71	5.86	1.78	6.17	1.84	6.46	1.90
0.7	5.83	1.86	6.15−	1.93	6.46	1.99	6.77	2.06
0.8	6.09	2.00	6.41	2.07	6.74	2.14	7.05+	2.21
0.9	6.33	2.13	6.67	2.21	7.00	2.28	7.32	2.35+
1.0	6.56	2.26	6.91	2.34	7.24	2.42	7.58	2.49
1.1	6.78	2.39	7.13	2.47	7.48	2.55−	7.82	2.63
1.2	7.00	2.51	7.36	2.60	7.71	2.68	8.06	2.76
1.3	7.20	2.64	7.57	2.72	7.93	2.81	8.28	2.89
1.4	7.40	2.76	7.77	2.85−	8.14	2.93	8.50+	3.02
1.5	7.60	2.88	7.98	2.97	8.35−	3.06	8.71	3.14
1.6	7.79	3.00	8.17	3.09	8.55−	3.18	8.92	3.26
1.7	7.97	3.11	8.36	3.20	8.74	3.30	9.12	3.39
1.8	8.15+	3.23	8.55−	3.32	8.94	3.42	9.32	3.51
1.9	8.33	3.34	8.73	3.44	9.12	3.53	9.51	3.62
2.0	8.50+	3.45+	8.91	3.55+	9.31	3.65−	9.70	3.74
2.1	8.67	3.56	9.09	3.67	9.49	3.76	9.89	3.86
2.2	8.84	3.68	9.26	3.78	9.67	3.88	10.07	3.97
2.3	9.01	3.79	9.43	3.89	9.84	3.99	10.25−	4.09
2.4	9.17	3.89	9.60	4.00	10.01	4.10	10.42	4.20
2.5	9.33	4.00	9.76	4.11	10.18	4.21	10.60	4.31
2.6	9.49	4.11	9.93	4.22	10.35+	4.32	10.77	4.43
2.7	9.65−	4.22	10.09	4.33	10.52	4.43	10.94	4.54
2.8	9.81	4.33	10.25−	4.44	10.68	4.54	11.10	4.65−
2.9	9.96	4.43	10.40	4.54	10.84	4.65+	11.27	4.76
3.0	10.11	4.54	10.56	4.65+	11.00	4.76	11.43	4.87

APPENDIX—TABLE I (*Concluded*)

λ	p_{16}	q_{16}	p_{17}	q_{17}	p_{18}	q_{18}	p_{19}
0.0	1.00	0.00	1.00	0.00	1.00	0.00	1.00
0.1	4.35+	0.91	4.56	0.94	4.77	0.98	4.97
0.2	5.07	1.19	5.31	1.23	5.54	1.28	5.77
0.3	5.60	1.42	5.85+	1.47	6.10	1.51	6.35+
0.4	6.04	1.62	6.30	1.67	6.57	1.72	6.83
0.5	6.42	1.80	6.70	1.85+	6.97	1.91	7.25−
0.6	6.76	1.97	7.05−	2.03	7.34	2.09	7.62
0.7	7.07	2.12	7.37	2.19	7.67	2.25−	7.96
0.8	7.37	2.28	7.68	2.34	7.98	2.41	8.28
0.9	7.64	2.42	7.96	2.49	8.27	2.56	8.58
1.0	7.91	2.56	8.23	2.64	8.55+	2.71	8.87
1.1	8.16	2.70	8.49	2.78	8.82	2.85−	9.14
1.2	8.40	2.84	8.74	2.91	9.07	2.99	9.40
1.3	8.63	2.97	8.97	3.05−	9.31	3.12	9.65−
1.4	8.86	3.10	9.21	3.18	9.55+	3.26	9.89
1.5	9.07	3.22	9.43	3.31	9.78	3.39	10.13
1.6	9.29	3.35−	9.65−	3.43	10.00	3.52	10.35+
1.7	9.49	3.47	9.86	3.56	10.22	3.64	10.58
1.8	9.69	3.59	10.07	3.68	10.43	3.77	10.79
1.9	9.89	3.71	10.27	3.80	10.64	3.89	11.00
2.0	10.09	3.83	10.47	3.92	10.84	4.01	11.21
2.1	10.28	3.95+	10.66	4.04	11.04	4.13	11.41
2.2	10.46	4.07	10.85+	4.16	11.23	4.25+	11.61
2.3	10.65−	4.18	11.04	4.28	11.43	4.37	11.81
2.4	10.83	4.30	11.22	4.39	11.61	4.49	12.00
2.5	11.00	4.41	11.41	4.51	11.80	4.61	12.19
2.6	11.18	4.53	11.58	4.62	11.98	4.72	12.38
2.7	11.35+	4.64	11.76	4.74	12.16	4.84	12.56
2.8	11.52	4.75−	11.94	4.85+	12.34	4.95−	12.74
2.9	11.69	4.86	12.11	4.96	12.52	5.06	12.92
3.0	11.86	4.97	12.28	5.07	12.69	5.18	13.10

APPENDIX—TABLE II

λ	p_0	q_0	p_1	q_1	p_2	q_2	p_3	q_3
0.00	0.0000		1.0000		1.0000		1.000	
0.01	0.0100	0.0000	1.0100	0.0000	1.0199	0.0000	1.039	0.000
0.02	0.0200	0.0100	1.0200	0.0100	1.0396	0.0198	1.077	0.039
0.03	0.0300	0.0200	1.0300	0.0200	1.0591	0.0392	1.114	0.075−
0.04	0.0400	0.0300	1.0400	0.0300	1.0785−	0.0583	1.150−	0.109
		0.0400		0.0400		0.0770		0.141
0.05	0.0500		1.0500		1.0976		1.184	
0.06	0.0600	0.0500	1.0600	0.0500	1.1166	0.0954	1.218	0.171
0.07	0.0700	0.0600	1.0700	0.0600	1.1354	0.1134	1.251	0.199
0.08	0.0800	0.0700	1.0800	0.0700	1.1541	0.1311	1.283	0.226
0.09	0.0900	0.0800	1.0900	0.0800	1.1726	0.1486.	1.314	0.252
		0.0900		0.0900		0.1658		0.276
0.10	0.1000		1.1000		1.1909		1.344	
0.11	0.1100	0.1000	1.1100	0.1000	1.2091	0.1826	1.374	0.300
0.12	0.1200	0.1100	1.1200	0.1100	1.2271	0.1993	1.403	0.322
0.13	0.1300	0.1200	1.1300	0.1200	1.2450+	0.2157	1.431	0.344
0.14	0.1400	0.1300	1.1400	0.1300	1.2628	0.2318	1.459	0.365−
		0.1400		0.1400		0.2477		0.384
0.15	0.1500		1.1500		1.2804		1.486	
0.16	0.1600	0.1500	1.1600	0.1500	1.2979	0.2634	1.513	0.404
0.17	0.1700	0.1600	1.1700	0.1600	1.3153	0.2789	1.539	0.422
0.18	0.1800	0.1700	1.1800	0.1700	1.3225+	0.2942	1.565−	0.440
0.19	0.1900	0.1800	1.1900	0.1800	1.3497	0.3093	1.590	0.458
		0.1900		0.1900		0.3242		0.475−
0.20	0.2000		1.2000		1.3667		1.615−	
0.21	0.2100	0.2000	1.2100	0.2000	1.3836	0.3389	1.639	0.491
0.22	0.2200	0.2100	1.2200	0.2100	1.4003	0.3534	1.663	0.508
0.23	0.2300	0.2200	1.2300	0.2200	1.4170	0.3678	1.687	0.523
0.24	0.2400	0.2300	1.2400	0.2300	1.4335+	0.3820	1.710	0.539
		0.2400		0.2400		0.3961		0.554
0.25	0.2500		1.2500		1.4500		1.733	
0.26	0.2600	0.2500	1.2600	0.2500	1.4663	0.4100	1.755+	0.569
0.27	0.2700	0.2600	1.2700	0.2600	1.4826	0.4238	1.778	0.583
0.28	0.2800	0.2700	1.2800	0.2700	1.4988	0.4374	1.800	0.597
0.29	0.2900	0.2800	1.2900	0.2800	1.5148	0.4509	1.821	0.611
		0.2900		0.2900		0.4643		0.625+
0.30	0.3000		1.3000		1.5308		1.843	
		0.3000		0.3000		0.4775+		0.639

APPENDIX—TABLE II (*Continued*)

λ	p_4	q_4	p_5	q_5	p_6	q_6	p_7	q_7
0.00	1.000	0.000	1.00	0.00	1.00	0.00	1.00	0.00
0.01	1.077	0.073	1.15−	0.13	1.26	0.21	1.43	0.29
0.02	1.147	0.135−	1.26	0.22	1.44	0.31	1.65+	0.36
0.03	1.212	0.188	1.37	0.28	1.58	0.36	1.81	0.39
0.04	1.272	0.233	1.46	0.33	1.69	0.40	1.92	0.42
0.05	1.329	0.274	1.53	0.37	1.78	0.43	2.02	0.44
0.06	1.382	0.309	1.61	0.40	1.86	0.45+	2.10	0.46
0.07	1.432	0.341	1.67	0.43	1.93	0.47	2.17	0.49
0.08	1.479	0.370	1.73	0.46	1.99	0.49	2.24	0.51
0.09	1.524	0.396	1.78	0.48	2.05+	0.51	2.30	0.53
0.10	1.567	0.420	1.84	0.50−	2.11	0.53	2.36	0.55+
0.11	1.608	0.443	1.88	0.52	2.16	0.55−	2.41	0.57
0.12	1.648	0.464	1.93	0.53	2.21	0.56	2.46	0.60
0.13	1.686	0.484	1.97	0.55+	2.25+	0.58	2.51	0.62
0.14	1.722	0.503	2.01	0.57	2.30	0.60	2.56	0.64
0.15	1.758	0.521	2.05+	0.58	2.34	0.62	2.60	0.66
0.16	1.792	0.538	2.09	0.60	2.38	0.63	2.65−	0.68
0.17	1.825+	0.555−	2.13	0.61	2.42	0.65+	2.69	0.70
0.18	1.857	0.571	2.16	0.63	2.46	0.67	2.73	0.72
0.19	1.889	0.586	2.20	0.64	2.49	0.68	2.77	0.74
0.20	1.919	0.601	2.23	0.66	2.53	0.70	2.80	0.75+
0.21	1.949	0.616	2.26	0.67	2.56	0.72	2.84	0.77
0.22	1.978	0.631	2.30	0.69	2.60	0.73	2.88	0.79
0.23	2.006	0.645−	2.33	0.70	2.63	0.75−	2.91	0.81
0.24	2.034	0.658	2.36	0.72	2.66	0.77	2.95−	0.82
0.25	2.061	0.672	2.39	0.73	2.69	0.78	2.98	0.84
0.26	2.088	0.685+	2.42	0.74	2.72	0.80	3.02	0.86
0.27	2.114	0.699	2.44	0.76	2.75+	0.81	3.05−	0.88
0.28	2.139	0.712	2.47	0.77	2.78	0.83	3.08	0.89
0.29	2.165−	0.724	2.50−	0.78	2.81	0.84	3.11	0.91
0.30	2.189	0.737	2.53	0.80	2.84	0.86	3.14	0.92

APPENDIX—TABLE II (*Continued*)

λ	p_8	q_8	p_9	q_9	p_{10}	q_{10}	p_{11}	q_{11}
0.00	1.00		1.00		1.00		1.00	
		0.00		0.00		0.00		0.00
0.01	1.63		1.83		2.01		2.16	
		0.33		0.32		0.30		0.30
0.02	1.87		2.07		2.24		2.40	
		0.36		0.35+		0.36		0.39
0.03	2.02		2.22		2.39		2.57	
		0.39		0.39		0.42		0.46
0.04	2.14		2.34		2.52		2.71	
		0.42		0.44		0.47		0.51
0.05	2.24		2.44		2.63		2.83	
		0.45−		0.48		0.52		0.56
0.06	2.32		2.53		2.73		2.93	
		0.48		0.51		0.56		0.60
0.07	2.40		2.61		2.82		3.03	
		0.51		0.55−		0.59		0.63
0.08	2.47		2.68		2.90		3.12	
		0.54		0.58		0.63		0.67
0.09	2.53		2.75+		2.98		3.20	
		0.56		0.61		0.66		0.70
0.10	2.59		2.82		3.05−		3.27	
		0.59		0.64		0.69		0.73
0.11	2.65−		2.88		3.11		3.34	
		0.62		0.67		0.71		0.75+
0.12	2.70		2.94		3.18		3.41	
		0.64		0.69		0.74		0.78
0.13	2.76		3.00		3.24		3.47	
		0.66		0.72		0.76		0.81
0.14	2.81		3.05+		3.29		3.53	
		0.69		0.74		0.79		0.83
0.15	2.85+		3.10		3.35−		3.59	
		0.71		0.76		0.81		0.85+
0.16	2.90		3.15+		3.40		3.65−	
		0.73		0.79		0.83		0.88
0.17	2.95−		3.20		3.45+		3.70	
		0.75+		0.81		0.86		0.90
0.18	2.99		3.25−		3.50+		3.75+	
		0.77		0.83		0.88		0.92
0.19	3.03		3.29		3.55+		3.80	
		0.79		0.85−		0.90		0.94
0.20	3.07		3.34		3.60		3.85+	
		0.81		0.87		0.92		0.97
0.21	3.11		3.38		3.64		3.90	
		0.83		0.89		0.94		0.99
0.22	3.15+		3.42		3.69		3.95−	
		0.85+		0.91		0.96		1.01
0.23	3.19		3.46		3.73		3.99	
		0.87		0.93		0.98		1.03
0.24	3.23		3.50+		3.77		4.04	
		0.89		0.94		1.00		1.05−
0.25	3.26		3.54		3.81		4.08	
		0.90		0.96		1.02		1.07
0.26	3.30		3.58		3.85+		4.12	
		0.92		0.98		1.03		1.09
0.27	3.34		3.62		3.89		4.16	
		0.94		1.00		1.05+		1.11
0.28	3.37		3.65+		3.93		4.20	
		0.96		1.02		1.07		1.12
0.29	3.40		3.69		3.97		4.24	
		0.97		1.03		1.09		1.14
0.30	3.44		3.72		4.01		4.28	
		0.99		1.05−		1.11		1.16

APPENDIX—TABLE II (*Continued*)

λ	p_{12}	q_{12}	p_{13}	q_{13}	p_{14}	q_{14}	p_{15}	q_{15}
0.00	1.00		1.00		1.00		1.00	
		0.00		0.00		0.00		0.00
0.01	2.30		2.44		2.59		2.74	
		0.32		0.36		0.40		0.43
0.02	2.56		2.73		2.90		3.07	
		0.43		0.47		0.49		0.51
0.03	2.75−		2.93		3.11		3.29	
		0.50−		0.53		0.55+		0.57
0.04	2.90		3.09		3.28		3.46	
		0.55+		0.58		0.60		0.63
0.05	3.03		3.22		3.42		3.61	
		0.60		0.62		0.65+		0.68
0.06	3.14		3.34		3.54		3.73	
		0.63		0.66		0.69		0.72
0.07	3.24		3.45−		3.65−		3.85−	
		0.67		0.70		0.73		0.76
0.08	3.33		3.54		3.75−		3.95+	
		0.70		0.73		0.77		0.80
0.09	3.42		3.63		3.84		4.05+	
		0.73		0.77		0.80		0.84
0.10	3.49		3.71		3.93		4.14	
		0.76		0.80		0.84		0.87
0.11	3.57		3.79		4.01		4.22	
		0.79		0.83		0.87		0.90
0.12	3.64		3.86		4.09		4.30	
		0.82		0.86		0.90		0.93
0.13	3.70		3.93		4.16		4.38	
		0.85−		0.89		0.93		0.96
0.14	3.77		4.00		4.23		4.45+	
		0.87		0.91		0.95+		0.99
0.15	3.83		4.06		4.29		4.52	
		0.90		0.94		0.98		1.02
0.16	3.89		4.13		4.36		4.59	
		0.92		0.96		1.01		1.05−
0.17	3.94		4.18		4.42		4.65+	
		0.94		0.99		1.03		1.07
0.18	4.00		4.24		4.48		4.72	
		0.97		1.01		1.06		1.10
0.19	4.05+		4.30		4.54		4.78	
		0.99		1.04		1.08		1.12
0.20	4.10		4.35+		4.59		4.83	
		1.01		1.06		1.10		1.15−
0.21	4.15+		4.40		4.65−		4.89	
		1.03		1.08		1.13		1.17
0.22	4.20		4.45+		4.70		4.95−	
		1.06		1.10		1.15+		1.20
0.23	4.25−		4.50+		4.75+		5.00+	
		1.08		1.13		1.17		1.22
0.24	4.30		4.55+		4.80		5.05+	
		1.10		1.15−		1.20		1.24
0.25	4.34		4.60		4.85+		5.10	
		1.12		1.17		1.22		1.26
0.26	4.39		4.65−		4.90		5.15+	
		1.14		1.19		1.24		1.29
0.27	4.43		4.69		4.95−		5.20	
		1.16		1.21		1.26		1.31
0.28	4.47		4.73		4.99		5.25+	
		1.18		1.23		1.28		1.33
0.29	4.51		4.78		5.04		5.30	
		1.20		1.25−		1.30		1.35−
0.30	4.55+		4.82		5.08		5.34	
		1.22		1.27		1.32		1.37

APPENDIX—TABLE II (*Concluded*)

λ	p_{16}	q_{16}	p_{17}	q_{17}	p_{18}	q_{18}	p_{19}
0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
0.01	2.90	0.45−	3.05+	0.46	3.20	0.47	3.35−
0.02	3.24	0.53	3.40	0.55−	3.56	0.57	3.72
0.03	3.46	0.60	3.64	0.62	3.81	0.65−	3.98
0.04	3.64	0.66	3.82	0.68	4.00	0.71	4.18
0.05	3.80	0.71	3.98	0.74	4.17	0.77	4.35+
0.06	3.93	0.75+	4.12	0.78	4.31	0.81	4.50−
0.07	4.05−	0.80	4.24	0.83	4.44	0.86	4.63
0.08	4.16	0.84	4.36	0.87	4.56	0.90	4.75+
0.09	4.26	0.87	4.46	0.91	4.67	0.94	4.87
0.10	4.35+	0.91	4.56	0.94	4.77	0.98	4.97
0.11	4.44	0.94	4.65+	0.98	4.86	1.01	5.07
0.12	4.52	0.97	4.74	1.01	4.95+	1.04	5.16
0.13	4.60	1.00	4.82	1.04	5.03	1.08	5.25−
0.14	4.68	1.03	4.90	1.07	5.12	1.11	5.33
0.15	4.75−	1.06	4.97	1.10	5.19	1.14	5.41
0.16	4.82	1.09	5.04	1.13	5.27	1.17	5.49
0.17	4.88	1.12	5.11	1.16	5.34	1.20	5.56
0.18	4.95−	1.14	5.18	1.18	5.41	1.22	5.63
0.19	5.01	1.17	5.24	1.21	5.48	1.25−	5.70
0.20	5.07	1.19	5.31	1.23	5.54	1.28	5.77
0.21	5.13	1.22	5.37	1.26	5.60	1.30	5.84
0.22	5.19	1.24	5.43	1.28	5.66	1.33	5.90
0.23	5.24	1.26	5.49	1.31	5.72	1.35+	5.96
0.24	5.30	1.29	5.54	1.33	5.78	1.38	6.02
0.25	5.35+	1.31	5.60	1.36	5.84	1.40	6.08
0.26	5.40	1.33	5.65−	1.38	5.89	1.42	6.14
0.27	5.45+	1.35+	5.70	1.40	5.95−	1.45−	6.19
0.28	5.50+	1.38	5.75+	1.42	6.00	1.47	6.25−
0.29	5.55+	1.40	5.80	1.45−	6.05+	1.49	6.30
0.30	5.60	1.42	5.85+	1.47	6.10	1.51	6.35+

APPENDIX—TABLE III

λ	p_0	q_0	p_1	q_1	p_2	q_2	p_3	q_3
0.000	.000	.000	1.000	.00000	1.000	.00000	1.000	.0000
0.001	.001	.001	1.001	.00100	1.002	.00199	1.004	.0040
0.002	.002	.002	1.002	.00200	1.004	.00399	1.008	.0079
0.003	.003	.003	1.003	.00300	1.006	.00599	1.012	.0119
0.004	.004	.004	1.004	.00400	1.008	.00797	1.016	.0158
0.005	.005	.005	1.005	.00500	1.010	.00995+	1.020	.0197
0.006	.006	.006	1.006	.00600	1.012	.01193	1.024	.0235+
0.007	.007	.007	1.007	.00700	1.014	.01390	1.028	.0273
0.008	.008	.008	1.008	.00800	1.016	.01587	1.032	.0311
0.009	.009	.009	1.009	.00900	1.018	.01784	1.035+	.0349
0.010	.010	.010	1.010	.01000	1.020	.01981	1.039	.0387
0.011	.011	.011	1.011	.01100	1.022	.02157	1.043	.0424
0.012	.012	.012	1.012	.01200	1.024	.02371	1.047	.0461
0.013	.013	.013	1.013	.01300	1.026	.02567	1.051	.0497
0.014	.014	.014	1.014	.01400	1.028	.02762	1.055−	.0534
0.015	.015	.015	1.015	.01500	1.030	.02955+	1.058	.0570
0.016	.016	.016	1.016	.01600	1.032	.03150+	1.062	.0606
0.017	.017	.017	1.017	.01700	1.034	.03343	1.066	.0642
0.018	.018	.018	1.018	.01800	1.036	.03537	1.070	.0678
0.019	.019	.019	1.019	.01900	1.038	.03729	1.074	.0713
0.020	.020	.020	1.020	.02000	1.040	.03922	1.077	.0748
0.021	.021	.021	1.021	.02100	1.042	.04114	1.081	.0783
0.022	.022	.022	1.022	.02200	1.044	.04306	1.085−	.0818
0.023	.023	.023	1.023	.02300	1.045+	.04498	1.089	.0852
0.024	.024	.024	1.024	.02400	1.047	.04688	1.092	.0886
0.025	.025	.025	1.025	.02500	1.049	.04880	1.096	.0920
0.026	.026	.026	1.026	.02600	1.051	.05070	1.100	.0954
0.027	.027	.027	1.027	.02700	1.053	.05260	1.103	.0988
0.028	.028	.028	1.028	.02800	1.055+	.05449	1.107	.1021
0.029	.029	.029	1.029	.02900	1.057	.05639	1.111	.1054
0.030	.030	.030	1.030	.03000	1.059	.05827	1.114	1.087

APPENDIX—TABLE III (*Continued*)

λ	p_4	q_4	p_5	q_5	p_6	q_6	p_7	q_7
0.000	1.000		1.000		1.00		1.00	
		.0000		.000		.000		.000
0.001	1.008		1.016		1.03		1.06	
		.0079		.016		.031		.058
0.002	1.016		1.031		1.06		1.12	
		.0157		.031		.058		.105+
0.003	1.024		1.046		1.09		1.17	
		.0233		.045+		.084		.144
0.004	1.031		1.061		1.12		1.21	
		.0309		.059		.107		.177
0.005	1.039		1.076		1.14		1.25+	
		.0382		.072		.128		.204
0.006	1.047		1.090		1.17		1.29	
		.0455−		.085−		.147		.227
0.007	1.054		1.104		1.19		1.33	
		.0526		.097		.165+		.247
0.008	1.062		1.118		1.22		1.36	
		.0596		.109		.182		.263
0.009	1.069		1.131		1.24		1.40	
		.0665−		.120		.197		.278
0.010	1.077		1.145−		1.26		1.43	
		.0732		.131		.211		.291
0.011	1.084		1.157		1.28		1.45+	
		.0799		.141		.224		.302
0.012	1.091		1.170		1.30		1.48	
		.0864		.151		.236		.311
0.013	1.098		1.183		1.32		1.51	
		.0928		.161		.247		.319
0.014	1.105+		1.195−		1.34		1.53	
		.0991		.170		.258		.327
0.015	1.112		1.207		1.36		1.55+	
		.1053		.179		.268		.334
0.016	1.119		1.219		1.37		1.58	
		.1115−		.188		.277		.341
0.017	1.126		1.231		1.39		1.60	
		.1175−		.197		.286		.346
0.018	1.133		1.242		1.41		1.62	
		.1234		.205−		.294		.351
0.019	1.140		1.253		1.42		1.63	
		.1292		.213		.302		.356
0.020	1.147		1.264		1.44		1.65+	
		.1349		.220		.309		.360
0.021	1.153		1.275+		1.45+		1.67	
		.1406		.228		.316		.365−
0.022	1.160		1.286		1.47		1.69	
		.1461		.235−		.322		.368
0.023	1.167		1.297		1.48		1.70	
		.1516		.242		.328		.372
0.024	1.173		1.307		1.50		1.72	
		.1570		.249		.334		.375+
0.025	1.180		1.317		1.51		1.74	
		.1623		.255+		.340		.379
0.026	1.186		1.328		1.52		1.75+	
		.1676		.261		.345−		.382
0.027	1.193		1.338		1.54		1.77	
		.1727		.268		.350−		.385−
0.028	1.199		1.347		1.55+		1.78	
		.1778		.274		.355−		.388
0.029	1.205+		1.357		1.56		1.79	
		.1828		.279		.359		.390
0.030	1.212		1.367		1.58		1.81	
		.1877		.285−		.364		.393

APPENDIX—TABLE III (*Continued*)

λ	p_8	q_8	p_9	q_9	p_{10}	q_{10}	p_{11}	q_{11}
0.000	1.00		1.00		1.00		1.00	
		.000		.000		.000		.000
0.001	1.12		1.21		1.35−		1.53	
		.104		.170		.240		.279
0.002	1.21		1.35+		1.53		1.72	
		.173		.246		.288		.278
0.003	1.29		1.46		1.65+		1.83	
		.220		.282		.297		.268
0.004	1.36		1.54		1.74		1.91	
		.253		.301		.296		.263
0.005	1.42		1.61		1.81		1.97	
		.277		.310		.295−		.265−
0.006	1.47		1.67		1.86		2.02	
		.295−		.315+		.294		.268
0.007	1.52		1.72		1.90		2.06	
		.308		.319		.295−		.275+
0.008	1.56		1.76		1.94		2.10	
		.318		.321		.296		.283
0.009	1.60		1.80		1.98		2.13	
		.325+		.323		.299		.291
0.010	1.63		1.83		2.01		2.16	
		.332		.324		.303		.300
0.011	1.66		1.86		2.04		2.19	
		.337		.326		.307		.310
0.012	1.69		1.89		2.07		2.22	
		.341		.329		.312		.319
0.013	1.72		1.92		2.09		2.24	
		.346		.331		.317		.328
0.014	1.74		1.94		2.12		2.27	
		.348		.333		.323		.338
0.015	1.77		1.97		2.14		2.29	
		.351		.336		.329		.347
0.016	1.79		1.99		2.16		2.31	
		.354		.339		.335−		.356
0.017	1.81		2.01		2.18		2.34	
		.356		.342		.341		.364
0.018	1.83		2.03		2.20		2.36	
		.359		.346		.347		.373
0.019	1.85+		2.05−		2.22		2.38	
		.361		.349		.353		.381
0.020	1.87		2.07		2.24		2.40	
		.364		.353		.360		.389
0.021	1.89		2.08		2.25+		2.42	
		.366		.357		.366		.397
0.022	1.91		2.10		2.27		2.44	
		.369		.360		.372		.405−
0.023	1.92		2.12		2.29		2.45+	
		.371		.364		.378		.412
0.024	1.94		2.13		2.30		2.47	
		.374		.368		.384		.420
0.025	1.95+		2.15−		2.32		2.49	
		.377		.372		.390		.427
0.026	1.97		2.16		2.34		2.51	
		.379		.377		.396		.434
0.027	1.98		2.18		2.35+		2.52	
		.382		.381		.402		.440
0.028	2.00		2.19		2.37		2.54	
		.384		.385−		.408		.447
0.029	2.01		2.20		2.38		2.55+	
		.387		.389		.414		.453
0.030	2.02		2:22		2.39		2.57	
		.390		.393		.419		.459

APPENDIX—TABLE III (*Continued*)

λ	p_{12}	q_{12}	p_{13}	q_{13}	p_{14}	q_{14}	p_{15}	q_{15}
0.000	1.00	.000	1.00	.000	1.00	.000	1.00	.000
0.001	1.71	.264	1.86	.220	1.98	.184	2.07	.175−
0.002	1.88	.239	2.01	.209	2.11	.207	2.21	.230
0.003	1.98	.253+	2.10	.225+	2.21	.243	2.32	.278
0.004	2.05−	.241	2.17	.247	2.28	.276	2.40	.316
0.005	2.10	.253	2.22	.270	2.34	.305−	2.47	.345−
0.006	2.15+	.267	2.27	.291	2.40	.329	2.54	.368
0.007	2.19	.281	2.32	.311	2.45+	.351	2.60	.387
0.008	2.23	.296	2.36	.329	2.50	.369	2.65+	.403
0.009	2.27	.310	2.40	.346	2.55−	.386	2.70	.417
0.010	2.30	.323	2.44	.362	2.59	.400	2.74	.429
0.011	2.33	.337	2.48	.376	2.63	.413	2.78	.440
0.012	2.36	.349	2.51	.389	2.66	.425−	2.82	.450−
0.013	2.39	.361	2.54	.401	2.70	.436	2.86	.459
0.014	2.42	.372	2.57	.412	2.73	.446	2.89	.468
0.015	2.44	.383	2.60	.423	2.76	.455−	2.93	.476
0.016	2.47	.393	2.63	.432	2.79	.463	2.96	.484
0.017	2.49	.403	2.65+	.442	2.82	.471	2.99	.491
0.018	2.52	.412	2.68	.450+	2.85−	.479	3.02	.499
0.019	2.54	.421	2.70	.459	2.87	.487	3.04	.506
0.020	2.56	.430	2.73	.467	2.90	.494	3.07	.513
0.021	2.58	.438	2.75+	.474	2.92	.501	3.09	.519
0.022	2.60	.446	2.77	.481	2.95−	.507	3.12	.526
0.023	2.62	.453	2.79	.488	2.97	.513	3.14	.532
0.024	2.64	.460	2.82	.495−	2.99	.520	3.16	.539
0.025	2.66	.467	2.84	.501	3.01	.526	3.19	.545−
0.026	2.68	.474	2.86	.507	3.03	.532	3.21	.551
0.027	2.70	.481	2.87	.514	3.05+	.537	3.23	.557
0.028	2.71	.487	2.89	.519	3.07	.543	3.25−	.563
0.029	2.73	.493	2.91	.525+	3.09	.549	3.27	.569
0.030	2.75−	.499	2.93	.531	3.11	.554	3.29	.575−

APPENDIX—TABLE III (*Concluded*)

λ	p_{16}	q_{16}	p_{17}	q_{17}	p_{18}	q_{18}	p_{19}
0.000	1.00	.000	1.00	.000	1.00	.000	1.00
0.001	2.16	.191	2.25−	.225+	2.35−	.267	2.46
0.002	2.32	.268	2.43	.308	2.56	.340	2.69
0.003	2.44	.319	2.57	.352	2.70	.372	2.84
0.004	2.53	.353	2.67	.379	2.81	.391	2.95+
0.005	2.61	.378	2.76	.397	2.90	.406	3.04
0.006	2.68	.396	2.83	.412	2.98	.419	3.12
0.007	2.75−	.412	2.90	.425−	3.04	.432	3.19
0.008	2.80	.425−	2.95+	.436	3.10	.444	3.24
0.009	2.85+	.436	3.01	.447	3.15+	.456	3.30
0.010	2.90	.446	3.05+	.457	3.20	.468	3.35−
0.011	2.94	.456	3.10	.467	3.25−	.479	3.40
0.012	2.98	.465+	3.14	.477	3.29	.490	3.44
0.013	3.02	.474	3.18	.486	3.33	.501	3.48
0.014	3.06	.482	3.21	.495+	3.37	.512	3.52
0.015	3.09	.491	3.25−	.505−	3.40	.522	3.56
0.016	3.12	.499	3.28	.514	3.44	.532	3.59
0.017	3.15+	.506	3.31	.522	3.47	.542	3.63
0.018	3.18	.514	3.34	.531	3.50	.551	3.66
0.019	3.21	.522	3.37	.539	3.53	.561	3.69
0.020	3.24	.529	3.40	.547	3.56	.569	3.72
0.021	3.26	.536	3.43	.556	3.59	.578	3.75−
0.022	3.29	.543	3.45+	.563	3.62	.587	3.78
0.023	3.31	.550+	3.48	.571	3.64	.595−	3.81
0.024	3.33	.557	3.50	.579	3.67	.603	3.83
0.025	3.36	.564	3.53	.586	3.69	.611	3.86
0.026	3.38	.571	3.55−	.593	3.72	.619	3.88
0.027	3.40	.577	3.57	.601	3.74	.626	3.91
0.028	3.42	.584	3.59	.608	3.76	.633	3.93
0.029	3.44	.590	3.61	.615−	3.78	.641	3.95+
0.030	3.46	.597	3.64	.621	3.81	.648	3.98

USE OF THE SIMPLEX DESIGN IN THE STUDY OF JOINT ACTION OF RELATED HORMONES

P. J. Claringbold

Department of Veterinary Physiology.
University of Sydney, N.S.W., Australia

INTRODUCTION

In certain toxicological studies the term joint action has come to take a special meaning. If members of a group of related compounds all cause death of an organism when administered separately the simultaneous action of these substances is called their joint action. Bliss (1939) first discussed the analysis of data obtained in this manner. From this time the problem has been examined in terms of tolerance distribution theory and developed in relation to probit analysis (Finney, 1952). Plackett and Hewlett (1951) have extended the tolerance distribution theory to different theoretical forms of joint action and developed a set of mathematical models, each of which is based on many assumptions and is very difficult to fit to experimental data.

Fisher (1954) has shown that parameters of the binomial distribution may be estimated without tolerance distribution assumptions. The aim of the present paper is to show that the study of joint action by means of an appropriate experimental design—the simplex design—allows ready interpretation of experimental data with no reference to a joint tolerance distribution, and no further assumptions than normally required in quantal analysis. The method is also appropriate without modification to the study of joint action of substances eliciting a graded response simply by applying the standard estimation procedures.

Examples will be drawn from the study of the action of oestrogens on the vagina of the ovariectomized mouse. The quantal response in this case is cornification of the vaginal epithelium.

MATHEMATICAL METHODS

The simplex design.

Suppose A_j, $(j = 1, 2, \cdots, k)$ are the doses of k hormones which, when administered separately, elicit approximately the same percentage response. A joint dose, D, may be defined in terms of k coordinates, X_j, which take positive values, thus,

$$D = \sum A_j X_j, \tag{1}$$

where the coordinate values are restricted by,

$$\sum X_i = 1. \tag{2}$$

The experimental region is therefore restricted by the above to a $(k - 1)$ dimensional simplex with vertices at the points on the coordinates $X_i = 1$. This method of approach allows all different types of joint doses to be uniquely specified. Thus if $X_i = 1$, the ith hormone is administered separately. If $X_i + X_j = 1$, then some mixture of the ith and jth hormones are administered together. In both the study of experimental designs in this region and the analysis of experimental data it is essential that a $(k - 1)$ dimensional coordinate system be introduced to the simplex. This may be done in two stages, (1) shift the origin of the X system to the centroid of the simplex, i.e. the point where every coordinate has the value $1/k$, (2) rotate the axes so that (say) the kth is orthogonal to the simplex.

The first is accomplished by the simple transformation (3) which at the same time changes the scale of measurement so that the vertices have non-fractional coordinates in the new system, $\overline{X}$.

$$\overline{X}_i = k(X_i - 1/k) = kX_i - 1, \tag{3}$$

where 1 is a vector of length k all elements of which are unity.

The second stage is carried out by an orthogonal transformation (4) of rank k with matrix, Θ. The scale is also modified so the vertex-centroid distance becomes $(k - 1)$ units.

$$\overline{\overline{X}}_i = k' \cdot \overline{X}_i \cdot \Theta \tag{4}$$

where $k' = k(k - 1) \cdot /k$ is a scale factor.

$$\Theta = 1/k(k - 1) \cdot \begin{bmatrix} k - 1 & 0 & 0 & \cdot & 0 & s \\ -1 & (k - 2)l & 0 & \cdot & 0 & s \\ -1 & -l & (k - 3)m & \cdot & 0 & s \\ -1 & -l & -m & \cdot & 0 & s \\ -1 & -l & -m & \cdot & 0 & s \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -l & -m & \cdot & n & s \\ -1 & -l & -m & \cdot & -n & s \end{bmatrix}$$

The additional letters in this matrix are determined from the fact that the sum of the squares of the elements of each column is $k(k - 1)$.

After this transformation all points of the simplex take the value 0 for the coordinate $\bar{X}_k$ which may therefore be ignored or used to describe another experimental variable, say different equivalent levels of dose. An experimental design consists of N points of the experimental region and may be summarized in a matrix called the *design matrix*, Box & Wilson (1951). The N rows of this matrix give the values of the coordinates at each of the N experimental points. In the present case the design matrix is of order N by $(k - 1)$.

An example when $k = 2$.

In this case equations 1 and 2 become

$$D = X_1 A_1 + X_2 A_2 , \quad X_1 + X_2 = 1, \quad 0 \le X_1 , \quad X_2 \le 1 \qquad (5)$$

The experimental region is a line. For illustrative purposes a design matrix consisting of 5 experimental points, including the vertices, the centroid and two intermediate points will be transformed using the appropriate forms of equations 3 and 4.

$$
\begin{array}{ccc}
X_1\ X_2 & \bar{X}_1\quad \bar{X}_2 & \bar{\bar{X}}_1\ \bar{\bar{X}}_2 \\[4pt]
\begin{bmatrix} 1 & 0 \\ \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{4} \\ 0 & 1 \end{bmatrix}
&
\begin{bmatrix} 1 & -1 \\ \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} \\ -1 & 1 \end{bmatrix}
&
\begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 0 \\ 0 & 0 \\ -\frac{1}{2} & 0 \\ -1 & 0 \end{bmatrix}
\end{array}
$$

where $\bar{X}_1 = 2X_1 - 1$, $\bar{X}_2 = 2X_2 - 1$.

$$[\bar{\bar{X}}_1 , \bar{\bar{X}}_2] = \tfrac{1}{2}[\bar{X}_1 , \bar{X}_2]\cdot\begin{bmatrix} 1, & 1 \\ -1, & 1 \end{bmatrix}$$

Transformation to a log dose scale.

Often in biological work response is linearly related to log dose. In studies on joint action it is of interest to test if this relationship still holds with respect to log joint dose. Since equation 1 is not linear in log joint dose a series of transformations are made so that this equation holds for the logarithms of the equivalent doses and the log joint dose in terms of different coordinates. These transformations are all based on the simple case $k = 2$.

Suppose equation 5 be written,

$$d = p a_1 + (1 - p)a_2 \qquad 0 \le p \le 1. \qquad (6)$$

This equation may be written,

if
$$\log d = q \log a_1 + (1 - q) \log a_2 ,\tag{7}$$

$$p = (r^{1-q} - r)/(1 - r), \qquad \text{where} \qquad r = a_2/a_1 .\tag{8}$$

Some values of this transformation are given in Table 1. Equation 7 may be put in the form of equation 5 by simple definition of terms, i.e. $D = \log d$, $A_1 = \log a_1$ and so on.

TABLE 1

Table of the transformation,

$$p = (r - r^{1-q})/(r - 1),$$

for equidistant sets of values of q. Geometric intervals of r are tabulated since the changes in p are more linear with this scale. (Figures in table are all $\times$ 10,000).

r

q	$\sqrt{2}$	2	$2\sqrt{2}$	4	$4\sqrt{2}$	8	$8\sqrt{2}$	16	32	64	128
1/2	5432	5858	6271	6667	7040	7388	7708	8000	8498	8889	9188
1/3	3725	4126	4531	4934	5330	5714	6083	6434	7071	7619	8079
2/3	7044	7401	7735	8042	8321	8571	8793	8987	9298	9524	9682
1/4	2834	3182	3541	3905	4271	4633	4988	5333	5983	6567	7082
3/4	7815	8108	8377	8619	8836	9026	9191	9333	9555	9710	9865
1/5	2286	2589	2904	3229	3558	3889	4217	4540	5161	5737	6260
2/5	4420	4843	5263	5675	6074	6454	6813	7148	7742	8234	8632
3/5	6410	6805	7180	7530	7853	8147	8411	8646	9032	9321	9530
4/5	8267	8513	8736	8935	9111	9263	9395	9506	9677	9794	9871
1/6	1916	2182	2461	2751	3047	3347	3648	3947	4529	5079	5589
5/6	8564	8775	8965	9134	9281	9408	9517	9608	9748	9841	9902
1/10	1163	1339	1528	1726	1933	2146	2363	2583	3023	3457	3875
3/10	3372	3755	4145	4537	4925	5304	5672	6024	6673	7241	7728
7/10	7354	7689	7998	8281	8536	8763	8962	9135	9410	9606	9741
9/10	9149	9282	9401	9514	9594	9670	9734	9787	9866	9918	9951

In more general cases the joint dose may be looked on as a series of equations 6. For example if $k = 3$ equation 1 may be written,

$$d = p\{p'a_1 + (1 - p')a_2\} + (1 - p)a_3$$

where $0 \leqq p$, $p' \leqq 1$. The quantity in braces may be regarded as a quantity, say b, and two transformations of the form of 8 made. Thus

$$d = a_1^{qq'} \cdot a_2^{q(1-q')} \cdot a_3^{(1-q)} .\tag{9}$$

where

$$p' = (r_1^{(1-q')} - r_1)/(1 - r_1), \qquad r_1 = a_2/a_1$$

$$p = (r_2^{(1-q)} - r_2)/(1 - r_2), \qquad r_2 = a_3/b.$$

Equation 9, on taking logarithms and using obvious definitions may be written,

$$D = X_1 A_1 + X_2 A_2 + X_3 A_3, \qquad X_1 + X_2 + X_2 = 1. \qquad (10)$$

In this form response may be related to log joint dose and its components in a simple manner. When equivalent doses are equal, logarithmic transformations need not be made, since in this case log joint dose is unaffected by variations in the coordinates subject to the restriction 2.

Extension of the experimental region.

Different equivalent levels of dose may be chosen for study. The method by which this is carried out depends on the form of the relationship of response to dose. In the case where this relationship is loglinear it is convenient to define each equivalent dose (A_i) as a function of an exponent (n) in terms of constants.

$$A_i = a_i r_i^n \qquad (11)$$

The values of the constants chosen depend on the Median effective dose (M.E.D.) and slope of the jth dose response line. Substituting the logarithm of these equations in equation 10 yields a function linear in n if the X's are held constant and linear in the X's if n is held constant. It is also useful to choose the levels of the constants so that the values of A_i chosen for study correspond to a set of equally spaced symmetric values of n centered at zero.

Other experimental variables may be introduced into the design in a factorial or other manner. In practice, however, if many points in the simplex are chosen for study this will lead to very large numbers of treatment combinations.

Analysis of variance.

Suppose a mixed level factorial experiment consists of the combinations of three factors denoted S, L and A at s, l and a levels respectively. The factor S is somewhat unusual and consists of s points of the simplex design, the factor L of l different levels of equivalent dose and A an additional factor at a levels. The complete design matrix therefore has $N = s \cdot l \cdot a$ rows and $(k - 1) + 2$ columns where k is the number of substances entering the simplex design. For each factor an orthogonal

set of comparisons including the identity may be drawn up and tabulated as an orthogonal matrix, or as the product of an orthogonal matrix and a diagonal matrix to preserve round numbers. Suppose these matrices be arranged so the columns present comparisons, the first column consisting only of unity i.e. the identity. Successive columns or comparisons may be numbered S^0, S^1, S^2, $\cdots$, where the superscript 0 denotes the identity and the other superscript has a currency of at most the number of degrees of freedom of the factor levels under consideration. The full set of orthogonal comparisons appropriate to the N treatment combinations may be obtained by the *direct product* (see Tocher, 1952 for a definition) of these matrices followed by an appropriate permutation of the columns. Since in general only main effects and first order interactions are required, other degrees of freedom going into an estimate or error or being isolated (see Fisher, 1951) only part of this product need be carried out. Main effect degree of freedom comparisons are obtained by the direct product of the column under consideration with all other identity columns. First order interaction comparisons are obtained by the direct product of the two individual main effect comparisons under consideration with the remaining identities. The matrix resulting from these direct products will consist of the first columns of an orthogonal matrix or the product of an orthogonal matrix with a diagonal matrix since the direct product of orthogonal matrices is orthogonal. The sum of squares attributable to the individual comparisons may be determined in the standard manner. For a binomial variable the appropriate procedures have been described by Claringbold, Biggers and Emmens (1953).

When $k = 2$, several sets of orthogonal comparisons have been determined for the purpose of detecting departures from linearity of response on dose. These are given for three cases, namely where one, two and three points are equally spaced on the line joining the two vertices.

Name

$S^{0'}$	1	1	1		1	1	1	1		1	1	1	1	1
$S^{1'}$	−1	0	1		−1	0	0	1		−1	0	0	0	1
$S^{2'}$	1	−2	1		−1	1	1	−1		1	0	−2	0	1
$S^{3'}$					0	−1	1	0		2	−3	2	−3	2
$S^{4'}$										0	−1	0	1	0

The first row (since the matrices have been transposed for convenience) is the identity. The second tests whether equivalent doses

were given. The third tests whether the mid-point response(s) falls on the line joining the control responses. The last comparison in each case is determined by those already made. An example of the use of these coefficients is given by Claringbold and Biggers (1955).

Sets of comparisons may be determined for other cases where k is greater than two and when symmetric arrays of points in the simplex have been chosen.

Regression analysis.

A response transformate may be directly related to functions of the coordinates of the design matrix by a weighted regression analysis. The information matrix in this case is not diagonal since the sums of squares and cross-products of the coordinates of the design matrix are not in general independent.

<center>EXAMPLE</center>

The data are summarised in Table 2 together with the coordinates of the experimental design. The plan of the simplex design used in this experiment is shown in Fig. 1. The complete design is in the form
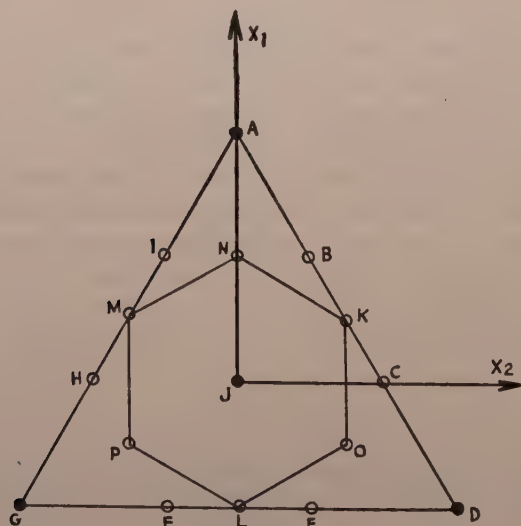
<center>FIG. 1.</center>

Plan of the two-dimensional simplex design used in the example. Points A, D and G correspond to the administration of oestrone, oestradiol-3:17β and oestriol alone, respectively. Points on the lines joining these vertices correspond to the administration of two oestrogen mixtures, while points within the triangle correspond to mixtures of three oestrogens.

of two equilateral triangular prisms, one for each replicate. Each prism has experimental points on three equidistant triangular planes. Since equal doses of oestrone, oestradiol-3:17β and oestriol are approximately equivalent in their effect on response when administered intravaginally no logarithmic transformations are used. The empirical angular response (Y) is related to ten functions of the coordinates of

TABLE 2

Percentage response of groups of 12 ovariectomized mice to joint intravaginal administration of oestrone, oestradiol and oestriol. The equivalent doses of these oestrogens denoted A_1, A_2, and A_3 were chosen so that

$$A_1 = A_2 = A_3 = 0.75 \times 10^{-4}\mu g. \quad \text{when } X_L = -1$$
$$= 1.50 \times 10^{-4}\mu g. \quad \text{when } X_L = 0$$
$$= 3.00 \times 10^{-4}\mu g. \quad \text{when } X_L = 1.$$

Original coordinates			Point	Coordinates in simplex		Response		
X_1	X_2	X_3		X_1	X_2	$X_L=-1$	$X_L=0$	$X_L=1$
First replicate—$X_R = -1$								
1	0	0	A	2	0	17	42	83
2/3	1/3	0	B	1	$t/3$	0	33	75
1/3	2/3	0	C	0	$2t/3$	33	33	75
0	1	0	D	−1	t	58	58	100
0	2/3	1/3	E	−1	$t/3$	17	33	67
0	1/3	2/3	F	−1	$-t/3$	33	33	58
0	0	1	G	−1	$-t$	25	50	42
1/3	0	2/3	H	0	$-2t/3$	25	42	42
2/3	0	1/3	I	1	$-t/3$	0	25	75
1/3	1/3	1/3	J	0	0	17	25	58
Second replicate—$X_R = 1$								
1	0	0	A	2	0	42	50	75
1/2	1/2	0	K	1/2	$t/2$	17	33	83
0	1	0	D	−1	t	75	67	83
0	1/2	1/2	L	−1	0	33	42	67
0	0	1	G	−1	$-t$	50	42	100
1/2	0	1/2	M	1/2	$-t/2$	17	42	58
2/3	1/6	1/6	N	1	0	33	33	58
1/6	2/3	1/6	O	−1/2	$t/2$	50	50	58
1/6	1/6	2/3	P	−1/2	$-t/2$	33	33	50
1/3	1/3	1/3	J	0	0	17	42	42

where $t = \sqrt{3}$

the design, by the following regression equation,

$$Y = \beta_0 + \beta_R X_R + \beta_1 \overline{\overline{X}}_1 + \beta_2 \overline{\overline{X}}_2 + \beta_L X_L + \beta_{12} \overline{\overline{X}}_1 \overline{\overline{X}}_2$$
$$+ \beta_{1L} \overline{\overline{X}}_1 X_L + \beta_{2L} \overline{\overline{X}}_2 X_L + \beta_{11} \overline{\overline{X}}_1^2 + \beta_{22} \overline{\overline{X}}_2^2 .$$

The information matrix was determined for these ten parameters and was inverted to give the variance-covariance matrix (Table 3). The

TABLE 3

Variance-covariance matrix for the experimental data and design given in Table 2. The theoretical variance used in its formation is that tabulated by Claringbold, Biggers and Emmens (1953) for the empirical angular transformation.

X_0	X_R	$\overline{\overline{X}}_1$	$\overline{\overline{X}}_2$	X_L	$\overline{\overline{X}}_1\overline{\overline{X}}_2$	$\overline{\overline{X}}_1 X_L$	$\overline{\overline{X}}_2 X_L$	$\overline{\overline{X}}_1^2$	$\overline{\overline{X}}_2^2$
3.281	−0.106	.	.	.	.	.	.	−1.066	−1.066
−0.106	1.261	.	.	.	.	.	.	0.056	0.056
.	.	3.182	.	.	.	.	.	−1.473	1.473
.	.	.	2.860	.	2.436	.	.	.	.
.	.	.	.	1.883	.	.	.	.	.
.	.	.	2.436	.	3.857	.	.	.	.
.	.	.	.	.	.	1.981	.	.	.
.	.	.	.	.	.	.	1.981	.	.
−1.066	0.056	−1.473	.	.	.	.	.	1.727	−0.605
−1.066	0.056	1.473	.	.	.	.	.	−0.605	1.727

matrix inversion was carried out using the method of Fox (1950) and Fox and Hayes (1952). In Table 4 the estimates of regression co-

TABLE 4

Regression analysis of the data of Table 2 following the empirical angular transformation.

Regression coefficient	Least square estimate	$t(\infty)$	P
β_0	35.39		
β_R	2.61 ± 1.12	2.3	$0.02 > P > 0.01$
β_1	1.93 ± 1.78	1.1	$0.3 > P > 0.2$
β_2	2.84 ± 1.69	1.7	$0.1 > P > 0.05$
β_L	12.30 ± 1.37	9.0	$P < 0.001$
β_{12}	-0.90 ± 1.96	0.5	$0.7 > P > 0.6$
β_{1L}	3.12 ± 1.41	2.2	$0.05 > P > 0.02$
β_{2L}	0.54 ± 1.41	0.4	$0.7 > P > 0.6$
β_{11}	3.20 ± 1.32	2.4	$0.02 > P > 0.01$
β_{22}	4.21 ± 1.32	3.2	$0.01 > P > 0.001$

Deviations from regression: $\chi^2_{(50)} = 49.7$, $0.7 > P > 0.5$.

efficients are tabulated together with their standard errors and test of significance. Both estimates of regression on the quadratic functions of the simplex coordinates are significantly positive. This indicates that the response to mixtures becomes smaller as the centroid of the simplex, which corresponds to a 1/3: 1/3: 1/3 mixture of the three oestrogens, is approached, and shows that the oestrogens have a mutually antagonistic action. The physiological significance of these findings is discussed by Claringbold (1955).

DISCUSSION

The simplex design in itself is a non-factorial design and may be criticised on these grounds. Factorial experiments in joint action studies lead to complex response surfaces even if one drug behaves simply as a dilution of the other (i.e. similar action, see Finney, 1952). Suppose a factorial experiment is designed for two factors (A, A') each at three levels. Suppose as a theoretical example both factors are simply doses of the one hormone, i.e., similar action must hold, and also suppose that response is linearly related to log dose. A possible design could be : —

		Dose of A (units)				Log_2 total dose		
		1	2	4				
Dose of A'	1	2	3	5		1.00	1.59	2.32
(units)	2	3	4	6		1.59	2.00	2.58
	4	5	6	8		2.32	2.58	3.00

The total dose administered to each animal in the nine groups of animals is shown in the body of the table, while the log total dose is shown as a subsidiary block of mixtures in one-one correspondence to the first block. If response is linear to log dose it must be proportional to these elements apart from some constant. Thus in the simplest case a curved response surface must be evaluated. Also if treatments consisting of one substance or control treatments are included they create difficulties since the log of zero is $-\infty$. The data must be analysed, therefore, in a number of disconnected steps. Plackett and Hewlett (1951) use this method and their analysis takes the following form:

1. Fit one substance dose response lines.
2. Predict on basis of alternative models the response to joint doses.
3. Choose the hypothesis which describes the observed data best.
Using the method described in this paper in the theoretical example,

the design could be:

α	Total dose (units)					
	1		2		4	
	Dose of A	Dose of A'	Dose of A	Dose of A'	Dose of A	Dose of A'
0	0	1	0	2	0	4
1/3	1/3	2/3	2/3	4/3	4/3	8/3
2/3	2/3	1/3	4/3	2/3	8/3	4/3
1	1	0	2	0	4	0

At each level of total dose administered four different methods of dividing the dose are shown. In the present example these must give equal responses and the response surface is easy to describe. Thus the present method has the advantages:

1. The one-substance treatments are simply special cases of the definition of a joint dose.

2. The data may be efficiently analysed in one step.

3. Similar action is indicated (in general) by no significant departures from linearity of response to log dose.

Finney (1952) uses an equation similar to equation 8 of this paper in the study of joint action. Instead of defining the transformation in terms of the actual doses administered it is defined in terms of relative potency, which is subject to estimation. If exactly equivalent doses were administered the transformation used here would be equivalent to that of Finney. Definition in terms of relative potency immediately restricts the study to the joint action of substances which give parallel dose response lines. Using the methods of this paper, series of approximately equivalent doses may be defined by appropriate geometric progressions and without any reference to relative potency. Claringbold and Biggers (1955) give an example where the joint administration of oestrone by two routes is studied. Here the slopes of the separate dose response lines are very different but the present method allowed ready interpretation of the response surface. In the present work although the mathematical equations do not demand equivalence of the doses administered this is desirable since the interval of linear relation of response to log dose is usually restricted. Thus although small departures from equivalence will not invalidate the present method large departures will lead to response surfaces difficult to evaluate.

I wish to thank Professor C. W. Emmens for advice during the course of this study.

REFERENCES

Bliss, C. I. (1939). The toxicity of poisons applied jointly. *Ann. Appl. Biol.*, *26*, 585–615.

Box, G. E. P. and Wilson, K. D. (1951). On the experimental attainment of optimum conditions. *J. R. statist. Soc.*, *13B*, 1–45.

Claringbold, P. J. (1955). Studies on the joint action of the natural oestrogens. In press.

Claringbold, P. J. and Biggers, J. D. (1955). The simultaneous administration of oestrone by the subcutaneous and intravaginal routes. In press.

Claringbold, P. J., Biggers, J. D. and Emmens, C. W. (1953). The angular transformation in quantal analysis. *Biometrics*, *9*, 467–484.

Finney, D. J. (1952). *Probit Analysis*. 2nd ed. Cambridge University Press.

Fisher, R. A. (1951). *The Design of Experiments*. 6th ed. Oliver and Boyd. Edinburgh.

Fisher, R. A. (1954). The analysis of variance with various binomial transformations. *Biometrics*, *10*, 130–151.

Fox, L. (1950). Practical methods for the solution of linear equations and the inversion of matrices. *J. R. statist. Soc.*, *12B*, 120–136.

Fox, L. and Hayes, J. G. (1951). More practical methods for the inversion of matrices. *J. R. statist. Soc.*, *13B*, 83–91.

Plackett, R. L. and Hewlett, P. S. (1952). Quantal responses to mixtures of poisons. *J. R. statist. Soc.*, *14B*, 141–163.

Tocher, K. D. (1952). The design and analysis of block experiments. *J. R. statist. Soc.*, *14B*, 45–100.

STATISTICAL ANALYSIS OF MULTIPLE SLOPE RATIO ASSAYS

C. G. Barraclough

Kraft Foods Limited, Melbourne, Australia

1. *Introduction.*

The statistical analysis of slope ratio assays for one test solution has been described in detail by Finney (1952), and routine methods of computation fully outlined, including tests for statistical and fundamental validity. Clarke (1952) has given a method for assays involving any number of test preparations, and this paper describes an adaptation of Clark's procedure using response totals directly to compute the various slopes, and gives a further analysis of the sum of squares associated with the test for fundamental validity. The method of analysis given here applies only to assays in which the response for each preparation is a linear function of the dose. The assay design must be completely symmetrical, i.e. there must be equal spacing between the dose levels for each preparation, the same number of dose levels for each preparation, and equal replication for all treatments. It is generally preferable to run a test at the zero dose level since this gives improved tests for validity, Finney (1952), Wood and Finney (1946); but the suggested method is developed to cover assays with, and without, tests at the zero dose level.

2. *Notation.*

An assay with a test at the zero dose level is termed an $(rk + 1)$ assay, while an assay without a zero dose level test is termed an (rk) assay.

Let x_{ij} = working dose, where $i = 1, 2, 3, \cdots r$, is the preparation and

$j = 0, 1, 2, \cdots k$ is the dose level for an $(rk + 1)$ assay

$j = 1, 2, \quad \cdots k$ is the dose level for an (rk) assay

The working scales are chosen so that the highest dose of each preparation is taken as unity, i.e. x_{ij} assumes the values $0, 1/k, 2/k, \cdots 1$, for each preparation in an $(rk + 1)$ assay.

If R'_i = potency estimate of the ith preparation on the working scales and

X_s = highest dose of standard preparation (taken as preparation 1)

X_{Ti} = highest dose of ith preparation (X_T is usually but not necessarily the same for each test preparation)

then

$$R_i = R'_i \times \frac{X_s}{X_{Ti}}$$

where R_i is the true potency estimate.

Let T_{ij} = response total for n replications of the dose x_{ij} , then

(1) $H_i = (2k - 2)T_{i1} + (2k - 5)T_{i2}$

$$+ (2k - 8)T_{i3} + \cdots - (k - 1)T_{ik}$$

where H_i is termed the intersection value for the ith preparation. The intersection value H_i is equal to the expected zero dose response total, multiplied by $[k(k - 1)]/2$, for the ith preparation, as estimated by a straight line fitted through the non zero dose response totals for that preparation.

The following symbols have been used for sums of squares and products for the doses and responses, to shorten the formulae.

(2) $$s_{11} = \sum_{i=1}^{r} \sum_{j=0}^{k} (x_{i_1 j} - \bar{x}_{i_1})^2$$

(3) $$s_{12} = \sum_{i=1}^{r} \sum_{j=0}^{k} (x_{i_1 j} - \bar{x}_{i_1})(x_{i_2 j} - \bar{x}_{i_2}) \qquad i_1 \neq i_2$$

(4) $$s_{i_1 T} = \sum_{i=1}^{r} \sum_{j=0}^{k} T_{ij}(x_{i_1 j} - \bar{x}_{i_1})$$

where $$\bar{x}_{i_1} = \frac{\displaystyle\sum_{j=0}^{k} x_{i_1 j}}{(rk + 1)} \qquad \text{for an } (rk + 1) \text{ assay}$$

and $$\bar{x}_{i_1} = \frac{\displaystyle\sum_{j=0}^{k} x_{i_1 j}}{rk} \qquad \text{for an } rk \text{ assay}$$

In the tests for validity orthogonal contrasts have been designated by

L_B = contrast associated with the test for "blanks"

L_{I_s} = contrasts associated with the tests for "intersection" where s runs from 1 to $(r - 1)$.

3. *General Formulae.*

The formal procedure for either the $(rk + 1)$ or (rk) type of assay involves the fitting of a multiple linear regression equation of the form

$$Y = a + b_1 x_1 + b_2 x_2 + \cdots + b_r x_r$$

where Y is the estimated mean response to the doses x_1, x_2, etc., of the various preparations, while b_1, b_2, etc. are the estimated increases in response per unit increase in dose of the corresponding preparations.

The potency estimate R'_i on the working scales is given by

$$R'_i = \frac{b_i}{b_1}$$

This is equivalent to fitting separate straight lines through the responses for each preparation, with the restriction that they all intersect at the zero dose level, and then obtaining the potency estimates from the ratios of the slopes of the lines for the test preparations to the slope of the line for the standard preparation. The formal method estimates the b_i values from the following type of equation

(5) $$b_i = v_{i1} S_{1T} + v_{i2} S_{2T} + \cdots + v_{ir} S_{rT}$$

where the v_{ii} are the elements of the variance and covariance matrix.

From the expression (4) for $S_{i,T}$ it can be seen that each term in (5) involves every T_{ij}, i.e. b_i can be expressed as a linear function of the T_{ij} values.

(6) $$b_{i_1} = \sum_{i=1}^{r} \sum_{j=0}^{k} m_{i_1 ij} T_{ij}$$

The values of the coefficients in (6) can be determined from (4) and (5) if the elements of the inverse matrix can be obtained in a convenient form. If we write

$$A = \begin{bmatrix} S_{11} & S_{12} & S_{12} & \cdots \\ S_{12} & S_{11} & S_{12} & \cdots \\ S_{12} & S_{12} & S_{11} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

where A is a square matrix of order r, then the variance and covariance matrix is given by

$$A^{-1} = M \begin{bmatrix} c & d & d & \cdots \\ d & c & d & \cdots \\ d & d & c & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

where

$$M = \frac{1}{(S_{11} - S_{12})(S_{11} + (r - 1)S_{12})}$$

$$c = S_{11} + (r - 2)S_{12}$$

$$d = -S_{12}$$

Since

$$S_{11} = \frac{n(k + 1)(2k + 1)}{6k} - \frac{n(k + 1)^2}{4(rk + 1)}$$

and

$$S_{12} = -\frac{n(k + 1)^2}{4(rk + 1)}$$

for an $(rk + 1)$ assay, where n is the number of replications, the values of M, c, d, can be readily calculated for any values of k and r. Table 4 contains values of M, c, d, for $r = 2$ to 10, and $k = 2, 3$ for an $(rk + 1)$ assay. They are used directly to determine the fiducial limits for R_i as discussed in the numerical example, and have been used in deriving the values of the multipliers. Table 5 contains the values for an (rk) assay obtained in the same way except that $(rk + 1)$ is replaced by (rk) in the expressions for S_{11} and S_{12} .

The values of M, c, d, could be combined with expression (5) to calculate the multipliers; but the symmetrical nature of the design permits (6) to be expressed more easily as

(7) $$b_i = m_0 T_0 + \sum_{j=1}^{k} m_j T_{ij} + \sum_{j=1}^{k} p_j \sum_{i=1}^{r} T_{ij}$$

T_0 is used for the total response for the n replications at the zero dose level since it is common to all preparations. The general form of (7) is the same for the two types of assay except that there is no term involving T_0 for an (rk) assay; but the multipliers have different values for the two types of assay. Values of the multipliers for an $(rk + 1)$ assay with $r = 2$ to 10, and $k = 2, 3$, are given in Table 6, while the multipliers for an (rk) assay are given in Table 7. A common factor

has been removed from each set of multipliers and is given in a separate column.

In routine assay work it is usually unnecessary to calculate the sum of squares due to regression since the regression will always be significant; but it is essential that validity tests be carried out on every assay. The most suitable tests are those for statistical and fundamental validity, generally referred to as 'blanks' and 'intersections' and discussed by Finney (1951) (1952), for the case of two preparations.

For r preparations the sum of squares in an analysis of variance associated with the test for blanks still has one degree of freedom but that for intersections has $(r - 1)$ degrees of freedom. The sum of squares for intersections can be further divided into $(r - 1)$ orthogonal contrasts, each with 1 degree of freedom, and these contrasts can be associated with specific tests among the intersection values for the various preparations. In a composite test for intersections it is possible that a significant result will be obtained, leading to the rejection of the whole analysis, when actually only one preparation is at fault. A suitable subdivision of the sum of squares for intersection would permit the isolation of the effect due to the faulty preparation, and if the remainder of the sum of squares for intersection was not significant the results of the assay could be recomputed to obtain valid results, after omitting the results for the faulty preparation. It is possible, though not very likely, that the composite sum of squares for intersection could give a non significant test, although one of the components would be significant if the subdivision was carried out. For (rk) assays there is no test for blanks; but the tests for intersections and the subdivision of the sum of squares for intersection can be carried out in the same manner as for an $(rk + 1)$ assay.

The r sums of squares are most conveniently obtained by using a table of orthogonal coefficients of the following form, in conjunction with the H values obtained from (4).

The orthogonality of the contrasts can be most clearly seen from the table but general formulae can be given for the contrasts.

$$(8) \qquad L_B = \frac{rk(k - 1)T_0}{2} - \sum_{i=1}^{r} H_i$$

$$\text{Divisor} = \frac{nrk(k - 1)\{2(2k + 1) + rk(k - 1)\}}{4}$$

$$(9) \qquad L_{I_s} = (r - s)H_s - \sum_{i=s+1}^{r} H_i \qquad\qquad s \text{ runs from 1 to } (r - 1).$$

$$\text{Divisor} = \frac{nk(k - 1)(2k + 1)(r - s + 1)(r - s)}{2}$$

Contrast	T_0	H_1	H_2	H_3	$\cdots$	H_{r-1}	H_r	Divisor
L_B	$\dfrac{rk(k-1)}{2}$	-1	-1	-1		-1	-1	$\dfrac{nrk(k-1)\{2(2k+1)+rk(k-1)\}}{4}$
L_{Γ_1}	0	$(r-1)$	-1	-1		-1	-1	$\dfrac{nk(k-1)(2k+1)r\cdot(r-1)}{2}$
L_{Γ_2}	0	0	$(r-2)$	-1		-1	-1	$\dfrac{nk(k-1)(2k+1)(r-1)(r-2)}{2}$
L_{Γ_3}	0	0	0	$(r-3)$		-1	-1	$\dfrac{nk(k-1)(2k+1)(r-2)(r-3)}{2}$
$\cdots$								
$L_{\Gamma_{r-1}}$	0	0	0	0		1	-1	$nk(k-1)(2k+1)$

The sum of squares associated with each degree of freedom =
L^2/divisor

If $k \geq 3$ the deviations from linearity for the non zero doses can be
calculated separately for each preparation using standard orthogonal
coefficients. For (rk) assays it is essential to have $k \geq 3$ so that tests
for deviations from linearity for the non zero doses can be carried out,
since there is no test for blanks. If the blanks component in an $(rk + 1)$
assay is significant; but the intersections component not significant, it
would be possible to discard the zero dose figures and analyse the re-
maining data as an (rk) assay, provided $k \geq 3$. It is suggested that
$k = 3$ is the best general purpose design since it has been shown by
Finney (1952), that the efficiency of a slope ratio assay falls rapidly
with increasing values of k.

4. *Numerical Example.*

The data used in the example are taken from the results of an assay
of niacin in yeast extracts. Five preparations were used, one standard
and four test, each at three levels, and a zero dose level test was in-
cluded. There were two replications, giving sixteen degrees of freedom
for the error estimate. The assay is based on the measurement of the
acidity produced by a culture of Lactobacillus arabinosus, Barton
Wright (1952), on a medium to which niacin has been added.

The figures given in Table 1 are the titres in mls. of $N/10$ sodium
hydroxide for each tube, while in Table 2 the duplicate measurements
have been totalled, and set out in a form more suitable for the compu-
tations.

TABLE 1

Preparation 1.		Preparation 2.		Preparation 3.	
0 μg.	3.2, 3.5	1 ml.	4.2, 4.7	1 ml.	3.8, 4.4
0.05 μg.	4.7, 4.8	2 ml.	5.0, 5.0	2 ml.	5.2, 5.4
0.10 μg.	6.2, 6.3	3 ml.	6.1, 6.1	3 ml.	6.2, 6.6
0.15 μg.	7.5, 7.7				
Preparation 4.		Preparation 5.			
1 ml.	4.0, 4.0	1 ml.	4.2, 4.3		
2 ml.	5.1, 5.6	2 ml.	5.2, 4.8		
3 ml.	6.0, 6.1	3 ml.	6.1, 6.3		

TABLE 2

	Prepn. 1 Standard	Prepn. 2	Prepn. 3	Prepn. 4	Prepn. 5	Totals
Zero Level	6.7					6.7
Level 1	9.5	8.9	8.2	8.0	8.6	43.2
Level 2	12.5	10.0	10.6	10.7	10.0	53.8
Level 3	15.2	12.2	12.8	12.1	12.4	64.7
b_i'	897.6	576.4	627.0	580.8	583.0	
R_i'		0.6422				
R_i		0.03211				
H_i	20.1	21.2	17.8	18.5	19.6	97.2

The multipliers necessary to calculate the slopes are obtained from
Table 6, for $r = 5$, $k = 3$. There is no need to use the common factor
given in Table 6, since we are only interested in the ratio of the slopes.
For this reason the slopes are denoted by b_i' instead of b_i .

$$b_1' = -42T_0 + 22T_{11} + 44T_{12} + 66T_{13}$$

$$- 24 \sum_{i=1}^{5} T_{i1} - 6 \sum_{i=1}^{5} T_{i2} + 12 \sum_{i=1}^{5} T_{i3}$$

$$= -(42 \times 6.7) + 22[9.5 + (2 \times 12.5) + (3 \times 15.2)]$$

$$- 6[(4 \times 43.2) + 53.8 - (2 \times 64.7)]$$

$$= 1762.2 - 864.6 = 897.\dot{6}$$

$$b_2' = 22[8.9 + (2 \times 10.0) + (3 \times 12.2)] - 864.6 = 576.4$$

$$\therefore R_2' = \frac{576.4}{897.6} = 0.6422 \qquad\qquad X_S = 0.15 \text{ μg. niacin}$$

$$X_T = 3 \text{ mls. test solution}$$

$$\therefore R_2 = 0.6\dot{4}22 \times \frac{0.15}{3} = 0.03211 \text{ μg. niacin per ml. test solution.}$$

The values of H_i are obtained from (4) as

$$H_i = 4T_{i1} + T_{i2} - 2T_{i3}$$

Thus $H_1 = (4 \times 9.5) + 12.5 - (2 \times 15.2) = 20.1$

A complete analysis of variance for the data is given in Table 3
but in routine assay work it is not necessary to compute the whole

TABLE 3

Analysis of Variance.

Source of Variance	Degrees of Freedom	Sums of Squares	Mean Squares
Between treatments	15	37.0650	2.4710
Regression	5	36.5429	7.3086
L_B	1	0.0165	0.0165
L_{I_1}	1	0.0130	0.0130
L_{I_2}	1	0.1176	0.1176
L_{I_3}	1	0.0248	0.0248
L_{I_4}	1	0.0144	0.0144
Q_1	1	0.0075	0.0075
Q_2	1	0.1008	0.1008
Q_3	1	0.0033	0.0033
Q_4	1	0.1408	0.1408
Q_5	1	0.0833	0.0833
Within treatments	16	0.7100	0.0444
Total	31	37.7750	

analysis. The essential parts are the error estimate obtained from the sum of squares within treatments, and the individual sums of squares for blanks, intersections, and deviations from linearity if $k \geq 3$. Using (8)

$$L_B = 15T_0 - \sum_{i=1}^{5} H_i = 100.5 - 97.2 = 3.3$$

Divisor $= 660$

∴ Sum of squares for blanks

$$= \frac{(3.3)^2}{660} = 0.0165$$

Using (9) we obtain

$$L_{I_1} = 4H_1 - \sum_{i=2}^{5} H_i = 80.4 - 77.1 = 3.3$$

Divisor $= 840$

∴ Sum of squares for the one degree of freedom corresponding to L_{I_1}

$$= \frac{(3.3)^2}{840} = 0.0130$$

TABLE 4

Elements of inverse matrix for $(rk + 1)$ assay.

r	k	M	c	d
2	2	$\dfrac{4}{35n}$	16	9
3	2	$\dfrac{4}{40n}$	17	9
4	2	$\dfrac{4}{45n}$	18	9
5	2	$\dfrac{4}{50n}$	19	9
6	2	$\dfrac{4}{55n}$	20	9
7	2	$\dfrac{4}{60n}$	21	9
8	2	$\dfrac{4}{65n}$	22	9
9	2	$\dfrac{4}{70n}$	23	9
10	2	$\dfrac{4}{75n}$	24	9
2	3	$\dfrac{9}{182n}$	31	18
3	3	$\dfrac{9}{224n}$	34	18
4	3	$\dfrac{9}{266n}$	37	18
5	3	$\dfrac{9}{308n}$	40	18
6	3	$\dfrac{9}{350n}$	43	18
7	3	$\dfrac{9}{392n}$	46	18
8	3	$\dfrac{9}{434n}$	49	18
9	3	$\dfrac{9}{476n}$	52	18
10	3	$\dfrac{9}{518n}$	55	18

TABLE 5
Elements of inverse matrix for (rk) assay.

r	k	M	c	d
2	2	$\dfrac{4}{10n}$	11	9
3	2	$\dfrac{4}{15n}$	12	9
4	2	$\dfrac{4}{20n}$	13	9
5	2	$\dfrac{4}{25n}$	14	9
6	2	$\dfrac{4}{30n}$	15	9
7	2	$\dfrac{4}{35n}$	16	9
8	2	$\dfrac{4}{40n}$	17	9
9	2	$\dfrac{4}{45n}$	18	9
10	2	$\dfrac{4}{50n}$	19	9
2	3	$\dfrac{9}{28n}$	8	6
3	3	$\dfrac{9}{42n}$	9	6
4	3	$\dfrac{9}{56n}$	10	6
5	3	$\dfrac{9}{70n}$	11	6
6	3	$\dfrac{9}{84n}$	12	6
7	3	$\dfrac{9}{98n}$	13	6
8	3	$\dfrac{9}{112n}$	14	6
9	3	$\dfrac{9}{126n}$	15	6
10	3	$\dfrac{9}{140n}$	16	6

L_{I_1} is a test of the average intersection value for the test solutions against the intersection value for the standard preparation. The other L_I contrasts are comparisons between the intersection values for the various test solutions. Since $k = 3$, only quadratic components of deviations from linearity for the non zero dose levels will exist.

Thus for Preparation I

$$Q_1 = \frac{(2T_{12} - T_{11} - T_{13})^2}{6n} = \frac{(0.3)^2}{12} = 0.0075$$

None of the mean squares for testing validity is significant so the data are consistent with the multiple regression equation, and the potency estimates are valid.

TABLE 6

Multipliers for an $(rk + 1)$ assay.

k	r	m_0	m_1	m_2	p_1	p_2	Common Factor
2	2	-15	7	14	-6	3	$\dfrac{2}{35n}$
2	3	-15	8	16	-6	3	$\dfrac{2}{40n}$
2	4	-15	9	18	-6	3	$\dfrac{2}{45n}$
2	5	-15	10	20	-6	3	$\dfrac{2}{50n}$
2	6	-15	11	22	-6	3	$\dfrac{2}{55n}$
2	7	-15	12	24	-6	3	$\dfrac{2}{60n}$
2	8	-15	13	26	-6	3	$\dfrac{2}{65n}$
2	9	-15	14	28	-6	3	$\dfrac{2}{70n}$
2	10	-15	15	30	-6	3	$\dfrac{2}{75n}$

TABLE 6—*Concluded*

Multipliers for an $(rk + 1)$ assay.

k	r	m_0	m_1	m_2	m_3	p_1	p_2	p_3	Common Factor
3	2	-42	13	26	39	-24	-6	12	$\dfrac{3}{182n}$
3	3	-42	16	32	48	-24	-6	12	$\dfrac{3}{224n}$
3	4	-42	19	38	57	-24	-6	12	$\dfrac{3}{266n}$
3	5	-42	22	44	66	-24	-6	12	$\dfrac{3}{308n}$
3	6	-42	25	50	75	-24	-6	12	$\dfrac{3}{350n}$
3	7	-42	28	56	84	-24	-6	12	$\dfrac{3}{392n}$
3	8	-42	31	62	93	-24	-6	12	$\dfrac{3}{434n}$
3	9	-42	34	68	102	-24	-6	12	$\dfrac{3}{476n}$
3	10	-42	37	74	111	-24	-6	12	$\dfrac{3}{518n}$

The fiducial limits for the various R values can be readily obtained using the approximate formula for the variance of R.

$$(10) \qquad V(R) = \frac{s^2 M}{b_1^2} \left(\frac{c}{X_T^2} - \frac{2Rd}{X_S X_T} + \frac{cR^2}{X_S^2} \right)$$

s^2 is the error mean square, while M, c, d are the appropriate values for the inverse matrix obtained from Table 4.

The approximate formula can be used provided g is less than 0.05 where

$$g = \frac{s^2 t^2 c M}{b_1^2} \qquad \text{and} \qquad t \text{ is Student's } t.$$

In most microbiological assays this condition will be satisfied; but if it is not satisfied the exact limits can be obtained from Fieller's formula,

(Fieller 1944). The expression (10) can be slightly simplified by using R' instead of R.

$$(11) \qquad V(R) = \frac{s^2 M}{b_1^2 X_T^2} [c\{1 + (R')^2\} - 2dR']$$

It is important to notice that the value of b_1 used in (10), (11) must be the correct one, not the b_1' given in Table 2.

5. *Acknowledgements.*

TABLE 7

Multipliers for an (rk) assay.

k	r	m_1	m_2	p_1	p_2	Common Factor
2	2	2	4	-6	3	$\dfrac{2}{10n}$
2	3	3	6	-6	3	$\dfrac{2}{15n}$
2	4	4	8	-6	3	$\dfrac{2}{20n}$
2	5	5	10	-6	3	$\dfrac{2}{25n}$
2	6	6	12	-6	3	$\dfrac{2}{30n}$
2	7	7	14	-6	3	$\dfrac{2}{35n}$
2	8	8	16	-6	3	$\dfrac{2}{40n}$
2	9	9	18	-6	3	$\dfrac{2}{45n}$
2	10	10	20	-6	3	$\dfrac{2}{50n}$

TABLE 7—*Concluded*

Multipliers for an (rk) assay.

k	r	m_1	m_2	m_3	p_1	p_2	p_3	Common Factor
3	2	2	4	6	-8	-2	4	$\dfrac{3}{28n}$
3	3	3	6	9	-8	-2	4	$\dfrac{3}{42n}$
3	4	4	8	12	-8	-2	4	$\dfrac{3}{56n}$
3	5	5	10	15	-8	-2	4	$\dfrac{3}{70n}$
3	6	6	12	18	-8	-2	4	$\dfrac{3}{84n}$
3	7	7	14	21	-8	-2	4	$\dfrac{3}{98n}$
3	8	8	16	24	-8	-2	4	$\dfrac{3}{112n}$
3	9	9	18	27	-8	-2	4	$\dfrac{3}{126n}$
3	10	10	20	30	-8	-2	4	$\dfrac{3}{140n}$

and assistance, and Miss M. Dick, Kraft Foods Limited, for supplying the data used in the example.

6. *References.*

Barton Wright, E. C. (1952). *The Microbiological Assay of the Vitamin B Complex and Amino Acids.* Sir Isaac Pitman & Sons Ltd., London.

Clarke, P. M. (1952). Statistical analysis of symmetrical slope ratio assays of any number of test preparations. *Biometrics, 8,* 377.

Fieller, E. C. (1944). A fundamental formula in the statistics of biological assay and some applications. *Quart. J. Pharm., 17,* 117.

Finney, D. J. (1951). The statistical analysis of slope ratio assays. *J. gen. Microbiol. 5,* 223.

Finney, D. J. (1952). *Statistical Method in Biological Assays.* Charles Griffin & Co., London.

Wood, E. C. and Finney, D. J. (1946). The design and analysis of microbiological assays. *Quart. J. Pharm. 19,* 112.

AN ANALYSIS OF PERENNIAL CROP DATA*

R. G. D. STEEL

Cornell University, Ithaca, N. Y.

SUMMARY

A bivariate analysis of variance is applied to perennial crop data. A test of an hypothesis about varietal effects is made. The bivariate analysis and a univariate analysis are compared. Two transformations of the data are considered. An expedient for locating varietal differences is proposed.

1. *Introduction*

In many fields of study, multiple observations are made on each individual. Treatment of such multivariate data differs. A trait-by-trait analysis may be made. Methods which consider several characters simultaneously include variance, covariance, components of variance and regression analysis. One of these may adequately answer the questions raised or test the hypotheses stated by the research worker when designing the experiment or survey. However, cases arise where none of these procedures is wholly adequate or appropriate. A multivariate analysis may be both appropriate and adequate in such cases. The term multivariate analysis will be applied to analyses of data where several variables are considered jointly with none relegated to the position of an independent variable.

If such multiple observations are analyzed on the basis of separate variables, the combination of the results of univariate tests and the assignment of a measure of credibility to any inference drawn present problems. Thus if the observations are perfectly correlated, the same conclusions are drawn from each variable; if the observations are completely independent and it is agreed to claim a difference at the 5% level if at least one variable shows significance, then one falsely claims significance with probability $1 - (.95)^n$ with n variables; if the rule is to claim a difference only if all variables show significance, then the

probability of falsely claiming a difference is $(.05)^n$ for n variables and it becomes practically impossible ever to detect a difference. In either case, rules can be constructed and inferences made with valid measures of credibility, but the true situation is probably somewhere between complete dependence and complete independence and we simply do not know the level of significance. In a multivariate analysis, the problem of dependence is looked after by the criterion itself.

It is the purpose of this paper to consider a multivariate analysis of yield for a forage crop where varieties are necessarily on the same plots in all years. The variables will be the total yields for each of two years. The tests of significance used are valid at the stated levels of significance whether or not year to year correlations exist or residual variances are homogeneous with respect to years.

Picture a graph, each point being the pair of variety means for the two years. If these paired means lie fairly closely together in a circle or ellipse, intuition suggests that they be declared not significantly different; whereas if they appear to lie along a line, extended rather excessively, intuition suggests variety differences that persist from year to year. For a 45° line, there appears to be no interaction of years and varieties whereas a line of other than 45° suggests a multiplicative effect of years, a special type of interaction not generally detected as such in the usual analysis of variance with years and its interactions as sources of variation. An additive year effect, solely or additionally, is indicated if the straight line is not through the origin. For any straight line, a single linear combination of the two years' yields should discriminate among varieties. The case where the points are scattered widely with little or no apparent linear correlation suggests a variety by year interaction other than the above special type. Discrimination here would require two linear functions.

A multivariate analysis will formally contain the ideas of the previous paragraph.

2. The Multi-variate Model

For a randomized complete block experiment, denote the yield in the h-th year for the i-th replicate and the j-th variety by

$$y_{ij}^{(h)} = \mu^{(h)} + \rho_i^{(h)} + \tau_j^{(h)} + \epsilon_{ij}^{(h)}.$$

Since tests of significance are planned, assume the $\epsilon_{ij}^{(h)}$'s have a joint normal distribution and for fixed h, are independently distributed with a common variance. Assumptions about the other additive components may be those of the usual models.

3. *The Data* and Analysis*

The data consist of the total plot-yields for each of 1949 and 1950 of 25 varieties of alfalfa planted in 1948 in a randomized complete block experiment with 4 replicates. The paired treatment means and overall mean for each variety are given in Table 1. The computations required

TABLE 1

Treatment means for 25 varieties of alfalfa in tons/acre, 4 replicates

Variety	1	2	3	4	5	6
1949	3.23	2.92	3.58	3.40	3.54	2.74
1950	4.47	4.25	4.15	4.52	4.66	3.80
Mean	3.85	3.58	3.87	3.96	4.10	3.27

7	8	9	10	11	12	13
2.78	3.14	3.53	3.51	3.44	3.68	3.18
4.10	3.76	4.55	4.58	4.02	4.86	4.26
3.44	3.45	4.04	4.04	3.73	4.27	3.72

14	15	16	17	18	19	20
3.62	3.28	3.68	3.54	3.46	3.28	3.37
4.06	4.07	3.71	4.44	4.26	3.84	3.91
3.84	3.68	3.69	3.99	3.86	3.56	3.64

21	22	23	24	25	Grand
2.94	3.16	3.58	3.40	3.44	3.34
4.04	3.87	4.52	3.86	3.83	4.18
3.49	3.51	4.05	3.63	3.63	3.76

initially are standard analyses of variance of the data for each year and the cross-products of an analysis of covariance. In Table 2, the 1949 and 1950 analyses are in the top left and lower right corners respectively, and the cross-products in the so-called off-diagonal. This presentation calls attention to the two-dimensional nature of the analysis.

*Data obtained through courtesy of C. C. Lowe, Department of Plant Breeding, Cornell.

Without reference to tests of significance, inferences must be qualitative. For such inferences, components of mean square (4, Tukey, 1949) may be helpful and are given. If one calculates correlation coefficients on the basis of components of mean squares, that for replicates is seen to be negative and greater than 1, whereas that for varieties is positive and of the order of .5.

TABLE 2
Bivariate analysis of variance

Source	d.f.	S. S.	M. S.	Component of M. S.
Replicates	3	$\begin{pmatrix} 4.2362 & -1.4819 \\ -1.4819 & 0.9074 \end{pmatrix}$	$\begin{pmatrix} 1.4121 & -0.4940 \\ -0.4940 & 0.3025 \end{pmatrix}$	$\begin{pmatrix} .0503 & -.0213 \\ -.0213 & .0049 \end{pmatrix}$
Varieties	24	$\begin{pmatrix} 6.9634 & 3.1027 \\ 3.1027 & 10.0438 \end{pmatrix}$	$\begin{pmatrix} 0.2901 & 0.1293 \\ 0.1293 & 0.4185 \end{pmatrix}$	$\begin{pmatrix} .0339 & .0225 \\ .0225 & .0595 \end{pmatrix}$
Residual	72	$\begin{pmatrix} 11.1171 & 2.8394 \\ 2.8394 & 12.9941 \end{pmatrix}$	$\begin{pmatrix} 0.1544 & 0.0394 \\ 0.0394 & 0.1805 \end{pmatrix}$	
Total	99	$\begin{pmatrix} 22.3167 & 4.4602 \\ 4.4602 & 23.9453 \end{pmatrix}$		

In order to assign a measure of uncertainty to an inference about varieties, let us test the null hypothesis that variety effects are zero for both $y^{(1)}$ and $y^{(2)}$. The criterion is

$$U = \frac{\begin{vmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{vmatrix}}{\begin{vmatrix} E_{11} + T_{11} & E_{12} + T_{12} \\ E_{21} + T_{21} & E_{22} + T_{22} \end{vmatrix}}$$

where E_{ij} and T_{ij} are sums of squares and cross-products for residuals and treatments respectively. Vertical bars indicate determinants are to be taken. The analogy between U and F is given by Tukey (4).

$0 \le U \le 1$ with values near one supporting the null hypothesis and values near zero indicating significant departures. The quantity $\sqrt{U}$ has been shown by Wilks (5, 6) to have a beta-distribution, with parameters p and q as used by Pearson (3) equal to (residual d.f. $- 1$) and

variety d.f. respectively. If it is desired to use F-tables, calculate

$$F = \frac{m}{n} \frac{1 - \sqrt{U}}{\sqrt{U}}$$

with $m = 2p$ and $n = 2q$ corresponding to d.f. for denominator and numerator respectively.

For this example

$$U = \frac{\begin{vmatrix} 11.1171 & 2.8394 \\ 2.8394 & 12.9941 \end{vmatrix}}{\begin{vmatrix} 18.0805 & 5.9421 \\ 5.9421 & 23.0379 \end{vmatrix}} = .357777$$

with $p = (72 - 1)$ and $q = 24$. Since

$$\sqrt{U} = .598, \qquad F = \frac{142}{48} \times \frac{.402}{.598} = 1.99$$

with m and $n = 142$ and 48 respectively. Variety effects are judged highly significant.

Significance raises the usual problems and the criterion U must be more closely considered. Reconsider this criterion in the determinantal equation

(1) $$\left| U\begin{pmatrix} E_{11} + T_{11} & E_{12} + T_{12} \\ E_{21} + T_{21} & E_{22} + T_{22} \end{pmatrix} - \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} \right| = 0.$$

This equation has two roots whose joint distribution (6) is

$f(U_1, U_2) \, dU_1 \, dU_2$

$$= K[(1 - U_1)(1 - U_2)]^{(n_1-3)/2} [U_1 U_2]^{(n_2-3)/2} (U_1 - U_2) \, dU_1 \, dU_2$$

where $1 \geq U_1 \geq U_2 \geq 0$, K is a known constant, and n_1 and n_2 are treatment and residual d.f. respectively. From this, the distribution of each root can be obtained and tested for significance.

The determinantal equation for the example is

$$381.2282U^2 - 457.3105U + 136.3945 = 0.$$

The roots are $U_1 = .6441$ and $U_2 = .5554$. Their joint distribution is

$f(U_1, U_2) \, dU_1 \, dU_2$

$$= K(1 - U_1)^{21/2}(1 - U_2)^{21/2} U_1^{69/2} U_2^{69/2}(U_1 - U_2) \, dU_1 \, dU_2 .$$

It is seen that the exact distributions of U_1 and U_2 can be obtained. Tukey (4) works an example for an odd and even pair of d.f. (It appears that he used location rather than variety d.f.)

We use χ^2 approximations. For testing U, it has kn_1 d.f. and is

$$\chi^2 = -[(n_1 + n_2) - \tfrac{1}{2}(k + n_1 + 1)] \log_e U$$

where k is the number of variates. Thus

$$\chi^2 = -[(24 + 72) - \tfrac{1}{2}(2 + 24 + 1)] \log_e .3578$$

$$= 90.95, \quad \text{d.f.} = 2 \times 24 = 48.$$

For testing U_1, it has $(k - 1)(n_1 - 1)$ d.f. and is

$$\chi^2 = -[(n_1 + n_2) - \tfrac{1}{2}(k + n_1 + 1)] \log_e U_1$$

We obtain $\chi^2 = 35.17$, d.f. $= 23$.

<center>TABLE 3</center>

Root	d.f.	χ^2
U_2	25	55.78**
U_1	23	35.17*
Total	48	90.95

Note that $U = U_1 \times U_2$ and that the smaller root, U_2, is tested first. The complement of U_2 is the square of a multiple correlation coefficient, which has been maximized by the choice of a linear combination of the two observations. A second linear combination, uncorrelated with the first, is associated with the complement of U_1. The new variates are *canonical variates*; the correlations are *canonical correlations*. They are further discussed in section 4.

We interpret the χ^2 table as follows: the significant value of U indicates real varietal differences. This suggests we examine U_2 and U_1. If U_1 were not significant, the significant U_2 would indicate that variety pairs did not depart significantly from a line, a space of one-dimension, and that varieties could be discriminated among by a single linear function of the paired yields. From the significant U_1, we conclude this is not the case; there appears to be variety $\times$ year interaction. The variety pairs fail to lie in a space of one dimension.

4. Transformations

In an analysis of variance with years as a source of variation, varieties and varieties $\times$ years mean squares are usually tested. These

involve a sum and a difference of the pairs of observations. Consider a linear transformation giving two new variables, multiples of the sum and difference, viz.,

$$(y_{ij}^{(1)}, y_{ij}^{(2)}) \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} = \left(\frac{y_{ij}^{(1)} + y_{ij}^{(2)}}{\sqrt{2}}, \frac{y_{ij}^{(1)} - y_{ij}^{(2)}}{\sqrt{2}} \right)$$

Multiplication is row-by-column. Note that the sum of the squares of the elements in each line of the transforming matrix add to unity and that the sum of the cross-products of the lines is zero. Such a matrix is said to be orthogonal. (If one had three years' data on a perennial crop, an appropriate transformation might involve the sum and linear and quadratic effects. Thus,

$$(y_{ij}^{(1)}, y_{ij}^{(2)}, y_{ij}^{(3)}) \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}$$

would be the transformation. The sum of squares of the elements in any line is unity and the sum of cross-products for any two lines is zero.)

In the univariate case if the variable x has variance s^2, then the variable ax has variance $a^2 s^2$. Analogously in the bivariate case, if the covariance matrix of $(y^{(1)}, y^{(2)})$ is

$$\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$$

then the covariance matrix of

$$(y^{(1)}, y^{(2)}) \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

is

(2)
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

$$= \begin{pmatrix} a^2 s_{11} + 2abs_{12} + b^2 s_{22} & acs_{11} + (ad + bc)s_{12} + bds_{22} \\ acs_{11} + (ad + bc)s_{12} + bds_{22} & c^2 s_{11} + 2cds_{12} + d^2 s_{22} \end{pmatrix}.$$

Applying (2) to the covariance matrices of Table 2, we obtain

TABLE 4

Bivariate analysis of $\left(\dfrac{\text{sum}}{\sqrt{2}}, \dfrac{\text{difference}}{\sqrt{2}} \right)$

Source	d.f.	M. Sq.	Components of M. Sq.
Replicates	3	$\begin{pmatrix} .3633 & .5548 \\ .5548 & 1.3513 \end{pmatrix}$	$\begin{pmatrix} .0063 & .0227 \\ .0227 & .0489 \end{pmatrix}$
Varieties	24	$\begin{pmatrix} .4836 & -.0642 \\ -.0642 & .2250 \end{pmatrix}$	$\begin{pmatrix} .0692 & -.0128 \\ -.0128 & .0242 \end{pmatrix}$
Residual	72	$\begin{pmatrix} .2068 & -.0130 \\ -.0130 & .1280 \end{pmatrix}$	

Components of mean square may be calculated directly from the new mean squares or by transforming the original components of mean square. The use of an orthogonal transformation matrix leaves unchanged the sum of the diagonal elements, or trace, of the covariance matrix. This serves as a partial check on the numerical results.

TABLE 5

Source	d.f.	S. S.	M. S.	Location
Years	1	35.0870	35.0870	——
Reps	3	1.0898	0.3633	Reps, left upper
Reps × Yrs	3	4.0538	1.3513	Reps, right lower
Varieties	24	11.6062	0.4836	Vars, left upper
Yrs × Vars	24	5.4010	0.2250	Vars, right lower
Reps × Vars	72	14.8952	0.2069	Residual, left upper
Residual	72	9.2160	0.1280	Residual, right lower
Total	199	81.3490		

The analysis of variance with years as a source of variation is given in Table 5. The column "Location" states where, in the bivariate analysis, the corresponding mean square is to be found. The mean square for years is available from the grand means of the bivariate

analysis as $(4.18 - 3.34)^2 100/2 = 35.28$, differing from that of the analysis of variance due to rounding of the means.

It is to be noted that only the diagonal terms of the bivariate analysis are in the analysis of variance. The bivariate analysis contains additional covariance terms often worth detecting by the experimenter.

The roots of the determinantal equation (1) remain unchanged by the transformation.

An alternate transformation suggested by the data themselves is available, viz. that which leads to canonical variates. The relation between the canonical variates first discussed by Hotelling (2) and those of this example is stated by Bartlett (1). For the first variate, the coefficients of $y^{(1)}$ and $y^{(2)}$ are found by solving the equations

$$(3) \qquad \left(R_1^2 \begin{pmatrix} E_{11} + T_{11} & E_{12} + T_{12} \\ E_{21} + T_{21} & E_{22} + T_{22} \end{pmatrix} - \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \right) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0$$

where $R_1^2 = 1 - U_2$. U_2 was the smaller root of equation (1). The complement of each root is a *canonical correlation*. Compare equations (2) and (1).

Bartlett (1) shows that this *canonical variate* is such that its treatment to treatment + residual sum of squares ratio, viz. R^2, is a maximum. Clearly the canonical variate is the *discriminant function* often defined as the linear function of the original observations for which the ratio of treatment to residual sum of squares is maximum.

Since both U_2 and U_1 are significant, the dependence of the data, after removal of replicate differences, upon variety effects is not adequately explained by the above canonical variate. A second canonical variate, uncorrelated with the first, may be obtained by replacing R_1^2 by $R_2^2 = 1 - U_1$ in equation (3).

For the first canonical variate, equation (3) is

$$\left((1 - .5554) \begin{pmatrix} 18.0805 & 5.9421 \\ 5.9421 & 23.0379 \end{pmatrix} - \begin{pmatrix} 6.9634 & 3.1027 \\ 3.1027 & 10.0438 \end{pmatrix} \right) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0$$

or $1.0752a_1 - .4608a_2 = 0$ and $-.4608a_1 + .1989a_2 = 0$. Hence $a_1 = .429a_2$ or $a_2 = 2.33a_1$ and the canonical variate may be written as $.429y^{(1)} + y^{(2)}$.

For the second canonical variate, equation (3) is

$$\left((1 - .6441) \begin{pmatrix} 18.0805 & 5.9421 \\ 5.9421 & 23.0379 \end{pmatrix} - \begin{pmatrix} 6.9634 & 3.1027 \\ 3.1027 & 10.0438 \end{pmatrix} \right) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0$$

and $a_1 = -1.868a_2$ or $a_2 = -.535a_1$. The canonical variate may be written as $1.868y^{(1)} - y^{(2)}$.

Dependence of the data upon variety effects is maximum for the variate $.429y^{(1)} + y^{(2)}$; any remaining dependence is maximum for $1.868y^{(1)} - y^{(2)}$. Neither variate seems particularly close to either $y^{(1)} + y^{(2)}$ or $y^{(1)} - y^{(2)}$, variates which have natural appeal. However, the fraction of (treatment + residual) sum of squares accounted for by variety effects is $R_1^2 = .4446$ for $.429y^{(1)} + y^{(2)}$ and is $24 \times .4836/(24 \times .4836 + 72 \times .2068) = .4380$ for $y^{(1)} + y^{(2)}$. For $1.868y^{(1)} - y^{(2)}$, the fraction is $R_2^2 = .3559$ and is $(24 \times .2250)/(24 \times .2250 + 72 \times .1280) = .3695$. The canonical variates are uncorrelated whereas the other two are not, though their correlation appears to be small. Apparently the variates $y^{(1)} + y^{(2)}$ and $y^{(1)} - y^{(2)}$, if appropriately used, could perform a satisfactory discrimination. (Of course, the variates to be considered depend upon the questions whose answers are required.) For a univariate model, there would seem to be little choice between one with additive year and variety effects and one with variety effects multiplied by a year constant.

The analysis of sums of squares for any new variate is

$$(a_1 , a_2)\begin{pmatrix} T_{11} + E_{11} & T_{12} + E_{12} \\ T_{21} + E_{21} & T_{22} + E_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$$= (a_1 , a_2)\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + (a_1 , a_2)\begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

When two canonical variates are required, as in this case, it may be desired to compute their bivariate analysis, vanishing of the correlation serving as a computational check. Using equation (2), we obtain Table 6. The transforming matrix is not orthogonal.

TABLE 6
Bivariate analysis of canonical variates

Source	d.f.	S. S.	M. S.
Replicates	3	$\begin{pmatrix} .4156 & .3549 \\ .3549 & 21.2257 \end{pmatrix}$	$\begin{pmatrix} .1385 & .1183 \\ .1183 & 7.0752 \end{pmatrix}$
Varieties	24	$\begin{pmatrix} 13.9875 & .0013 \\ .0013 & 22.7504 \end{pmatrix}$	$\begin{pmatrix} .5828 & .0001 \\ .0001 & .9479 \end{pmatrix}$
Residual	72	$\begin{pmatrix} 17.4763 & .0007 \\ .0007 & 41.1784 \end{pmatrix}$	$\begin{pmatrix} .2427 & .0000 \\ .0000 & .5719 \end{pmatrix}$

TABLE 7

Variety	1	2	3	4	5
$.429y^{(1)} + y^{(2)}$	5.85	5.50	5.69	5.98	6.18
$1.868y^{(1)} - y^{(2)}$	1.56	1.20	2.55	1.84	1.95

6	7	8	9	10	11	12
4.98	5.30	5.10	6.06	6.08	5.50	6.44
1.33	1.09	2.11	2.05	1.98	2.40	2.02

13	14	15	16	17	18	19
5.63	5.62	5.48	5.29	5.96	5.74	5.25
1.68	2.71	2.06	3.15	2.18	2.22	2.28

20	21	22	23	24	25	Grand
5.36	5.30	5.22	6.06	5.32	5.30	5.61
2.39	1.45	2.04	2.16	2.49	2.59	2.06

5. *The Canonical Variates and the Interpretation of the Data*

Table 7 contains the values of the 25 transformed means. These variables are not correlated and there is little value in observing their graph. This property of independence is useful in making exact probability statements.

Interpretation of the data comes within the province of the experimenter. The preceding analysis, indicating the need of two discriminant functions, raises an even more difficult problem than does a significant F in a univariate analysis. This was to be expected since a multivariate analysis broadens the basis of our null hypothesis and requires no assumption about homogeneity of variance, an assumption that may be false.

As a temporary expedient, the author proposes the following analysis of variance technique. From Table 6, obtain the residual sum of squares and divide by $(72 - 1) \times 4$ to obtain the variance of a treatment mean for 4 replicates. The use of $(72 - 1)$ in place of 72 is due to

the fact that the ratio of the coefficients for the first canonical variate was obtained from the data. We obtain $s_{\hat{z}}^2 = .0616$ and $s_{\hat{z}} = .25$. Discriminate among means for the first canonical variate (Table 7) by some standard technique such as Duncan's New Multiple Range Test using (number of means $+$ 1) where the number of means is required. Justify this on the basis that the new variate has a ratio of coefficients determined from the data. The use of (residual d.f. $-$ 1) and (number of variates $+$ 1) is suggested by the analysis of variance argument often presented with a discriminant function analysis.

For a perennial crop at a single location, the experimenter is presumably interested in a variety which, for yield, is persistently good. Such varieties can be discriminated among by a single function. Thus, an analysis of the second canonical variate, perhaps as above using (residual d.f. $-$ 2), should be with a view to finding out something about the variety and/or year that produced significance. This analysis helps locate varieties that are consistently good (poor) but sometimes do even better (poorer) than expected and ones that are good (poor) in some years and not exceptional or are even poor (good) in others. Other characteristics of such varieties or the frequency of the sort of year, i.e. the total environment, that produced such results should lead to a decision on retaining or discarding such varieties.

REFERENCES

(1) Bartlett, M. S. Multivariate analysis. *Jour. Roy. Stat. Soc. Suppl.*, *9*, 1947, 176–197.

(2) Hotelling, H. Relations between two sets of variates. *Biometrika*, *28*, 1936, 321–377.

(3) Pearson, K. *Tables of the incomplete beta function.* Cambridge University Press, 1932.

(4) Tukey, J. W. Dyadic anova, an analysis of variance for vectors. *Human Biology*, *21*, 1949, 65–110.

(5) Wilks, S. S. Test criteria for statistical hypotheses involving several variables. *Jour. Am. Stat. Ass'n.*, *30*, 1935, 549–560.

(6) Wilks, S. S. *Mathematical Statistics.* Princeton University Press, 1946.

AN INTRODUCTORY COURSE IN BIOMETRY FOR GRADUATE STUDENTS IN BIOLOGY*

C. I. BLISS

*The Connecticut Agricultural Experiment Station and
Yale University, New Haven, Connecticut*

At the Third International Biometric Conference in 1953, a symposium on the first course in biometry was organized by Professor W. G. Cochran, who provided each participant with a list of topics for discussion. The present paper, from this symposium, is based upon eleven years of teaching biometry to graduate students in biology at Yale University. These students are potential research biologists and the course is intended to provide them with an essential research tool. During this period, 86 students have received University credit for the course and perhaps 25 more were serious auditors. Since these students would judge the effectiveness of the course from a different viewpoint than their instructor, I both queried my last two classes and sent a questionnaire to all earlier students whose address was known. Of some 85 questionnaires distributed, 75 have been returned. These student opinions will be considered in relation to each topic on the agenda of the symposium.

Interests and preparation of the instructor and students. Although biometry concerns both the mathematical and statistical aspects of biology, the content of an introductory course depends upon the interests and preparation of both the instructor and the students. My course has been primarily statistical. This was unavoidable in view of the limited mathematical background of the majority of my students. It may be rationalized by the readier applicability of statistics than of mathematics in most biological research. The mathematical models that suffice for the statistical aspects are relatively simple, and can be applied in areas as distinct as botany, pharmacology, zoology, forestry, microbiology and the medical and agricultural sciences. Graduate students from most of these fields have attended my course, often in the same class.

Biometry can be defined in so many ways that the viewpoint of the instructor largely determines the character of a course. Hence, it is pertinent to report my own background, which was primarily biological, starting with undergraduate and graduate majors in zoology and

followed by seven years as a research entomologist. My research projects required an increasing use of statistical method, so that later I studied for two years with Professor R. A. Fisher, and since 1938 have worked entirely as a biometrician, primarily on experimental problems in agriculture and pharmacology. In 1943, I began teaching biometry to graduate students at Yale University, originally in two alternating courses, one primarily for pharmacologists and the other for botanists and zoologists. These were combined in 1950. Despite many changes in content and approach through the years, all students will be assumed here to have taken a single course.

The distribution of students among the different biological fields is shown in Table 1, about 90 percent of them being men. Only two of

TABLE 1

Major field of students in the course and field of employment, where known, of graduates.

Field	Number in each field as		Number of graduates
	Students	Auditors	
Pharmacology	31	10	28
Other medical sciences	9	6	11
Zoology	14	1	10
Forestry	21	2	13
Other plant sciences	9	2	8
Mathematics and statistics	2	3	5
Other areas	—	1	7
Total	86	25	82

those taking the course for credit had majored in either mathematics or statistics. The majority may have had introductory calculus, usually so long before that it had been largely forgotten. To insure a common basis, an initial chapter in the Outline, which now serves as our text, reviews the elementary mathematics that is assumed. If any of it seems strange, the student is referred to the book by Professor Walker (6). Although statistics is not a prerequisite, about one student in five has taken the subject before, and a few more have had lectures on statistics in other courses. Since completing their biometry at Yale, about one in ten has taken further work in statistics.

The course has not been a recruiting ground for professional statisticians or biometricians, as evidenced in Table 1 by the employment, where known, of the students who have graduated. By and large, each has remained in biology, in the field for which he trained;

only three or four have gone into statistics or biometry professionally. So far as their activities could be determined, about 58 percent are primarily in research, about 30 percent in teaching and the remaining 12 percent in other activities, usually neither biometrical nor statistical.

Purpose of the course. The objective of the course is to train future professional biologists to use intelligently the statistical methods required in the design and analysis of biological investigations. How well was this purpose achieved?

One criterion would be the extent to which they applied statistical techniques after graduation. The questionnaire asked how extensively, on a scale of 0 to 3, each respondent was involved in the design and in the analysis of experiments, and on the same scale, how much biometry was used in the process. Four out of five respondents gave scores of 1 to 3 to this kind of activity, and within this range their scores averaged 2.0 and 2.1 for design and analysis respectively. The more these former students were involved in either operation, the more they utilized their biometry ($P \doteq .01$).

Another question asked them to list the principal statistical techniques that they had used since taking the course. As summarized in Table 2, the list includes methods that would be considered fairly sophisticated. Some of them have not been taught until recently, so that the relative frequencies are only suggestive. These techniques were listed under three headings: (a) "methods where the information gained from my course was sufficient"—totalling 64 percent, (b) "methods where a moderate amount of additional study enabled you to proceed on your own"—22 percent, and (c) "methods covered so briefly, if at all, that you had to learn them almost entirely from other sources"—14 percent. The extent to which a statistical method was used depended in large part upon its having been included in the course. Relatively few were learned later *de novo*. This seems to me a good reason for covering much ground rapidly rather than less more thoroughly.

A student should not complete a course with a false idea of how much he knows but be more apt than before to consult a trained statistician. One question, therefore, read as follows: "Have you consulted a statistician or biometrician in connection with your research? If so, to what extent has my course prepared you for these conferences?" Of 65 who answered the first question, 36 had consulted a statistician, and of the 35 who answered the second question, 32 considered the preparation given by the course as adequate, the other three giving qualified answers, such as "fine, but only after considerable experience". The following are some comments: "I think this is the chief value of a

TABLE 2

Statistical techniques used after completing the course and the frequency with which each was mentioned in 75 questionnaires.

Technique	Frequency
Analysis of variance	39
Tests of significance, t, F, χ^2	31
Experimental design, randomized blocks, Latin squares	27
All-or-none bioassays, LD50's, probit analysis	22
Regression	21
Bioassays with a graded response	17
Analysis of covariance	16
Correlation	15
Factorial design and analysis	12
Partial regression	7
Poisson distribution	6
Sampling techniques	5
Standard deviation, standard error, etc.	5
Components of variance	4
Binomial probabilities	3
Comparison of means	3
Multiple assays	3
Negative binomial	3
Partitioning hereditary components	3
Slope-ratio assays	3
6 techniques*	2
10 techniques**	1

*Discriminant function, tests for normality, estimation of number of observations needed, quality control, mathematical derivations, construction of mathematical and probability models.

**Effects of population density, transformations, antagonism and synergism, confidence limits, power functions, non-normal distributions, practical mathematical statistics, expectation, test construction and validation, graphic representation of equations.

good course in statistics. Most experiments do not lend themselves to routine treatment and a well-trained statistician needs to be consulted intelligently." "New problems readily recognized." "Course has been indispensable for understanding biometricians." These and other replies indicate a lively appreciation of the value of the statistical consultant in biological research.

Time required. Most students find biometry a difficult subject and students and faculty alike have protested the time it requires, representing about 1/8 of the course work for a doctor's degree. To a biometrician this is not excessive in view of the basic importance of the subject, but to a subject-matter department, the requirement seems alarmingly high. In trying to meet these complaints, I have experi-

mented constantly in my teaching, so that the course has changed over the years in content, in timing, and in many other details.

Initially, I gave a weekly two-hour lecture, plus laboratory, for one 14-week semester. Most students seemed to reach the saturation point after one hour, so that we shifted to two one-hour lectures, extending the course to two semesters. This conflicted with one of the graduate programs, so that we changed again to three one-hour lectures a week for one semester, and in alternate years I gave the hardier students two one-hour lectures a week in the second semester. After a considerable trial, three lectures a week has proved too concentrated, and we are now returning to two lectures a week through the academic year.

In our experience, most statistics is learned by working illustrative examples and studying their meaning, so that the statistical laboratory has been a basic part of our program from the very start. Nominally, two hours of guided laboratory instruction is required for each hour of lecture, but students usually need an additional two hours of laboratory on their own.

This past year the small size of my class has enabled me to give each student an individual 45-minute tutorial each week, in addition to lectures and laboratory. When queried at the end of term on procedure for a larger class, one would prefer a session alone in alternate weeks, two a weekly session with several students, and one tutorials held in the laboratory. All of them urged continuing the experiment.

My attitude towards examinations has changed through the years. At the beginning each student was assigned a problem at the end of the course and asked to turn in an answer at his convenience. Now I give several examinations in a semester, each consisting of a closed-book, written quiz, and of one or more problems to be worked on the calculator with books open. Restricting each examination to the material covered in one section of the course has improved student morale.

Statistical Laboratory. Except for an introductory taste test, the laboratory exercises consist primarily of computing and understanding selected numerical examples. After working through a variety of applications, the student should recognize more readily opportunities for increasing the efficiency of his own research. A laboratory assistant, who has majored or minored in mathematics, holds individual or group conferences, quizzing each student on what he is doing and why.

To allow some selection, the examples number about 250 in the first 17 chapters of the syllabus but students seldom work as many as one in five. An electric calculator is provided, and tedium is reduced by supplying the basic terms for each problem wherever possible, such as

the total sum of squares. Very few examples have been invented or reduced artificially to one or two digits. The solution of full-sized examples is intended to reduce the gap between the form in which raw data reach the investigator and the form to which the student becomes accustomed. Several students in fact have recommended that some examples should be completely unclassified, with no clues beyond a statement of the biologist's objective. One called for "more experience in handling raw data, following a work form is not sufficient", a recommendation that has been strongly seconded by my most recent class.

The importance of the laboratory cannot be overestimated. Auditors without a statistical background who skip the laboratory soon find the lectures unintelligible. During lecture, students are introduced to the subject, but they learn it only in the laboratory. One question in the questionnaire was: "How would you rate the relative usefulness of lectures and laboratory?" Of the 62 replies, 26 considered the laboratory more useful than the lectures, 21 thought they were of equal value and seven found the laboratory less useful than the lectures. Their comments were often emphatic, among them the following: "I did my actual learning in the laboratory but lecture discussions are an essential supplement." "A course such as this demands both laboratory and lecture, one without the other would be of little value." "Significance of what was said in the lecture often didn't strike home until after a few specific problems were wrestled with." "That material on which I had spent most time in laboratory has stayed with me and has been more useful." "Lectures more useful in the latter part of the course after we had learned the basic principles of statistics." "Need constant practice in the laboratory to grasp lecture material."

Course content. The course is taught from an Outline which has been developed over a period of years (1). Its primary purpose is to free the student of basic note-taking during lecture, but it is sufficiently detailed to serve as the principal text. Readings are assigned in Fisher's "Statistical Methods for Research Workers" (2) and "The Design of Experiments" (3), and students are required to have "Statistical Tables for Biological, Agricultural and Medical Research" by Fisher and Yates (4). While it was in preparation, lectures followed the Outline very closely.

The course starts with a class experiment based upon the tea tasting test in Fisher's "Design of Experiments" but substituting fresh and reconstituted skim milk. This leads to the binomial distribution, the χ^2 distribution and contingency tables. Following chapters on the normal distribution, and interval estimation, the class has its first examination. The analysis of variance is introduced with the comparison

of two groups and developed progressively throughout the remainder of the course. The chapter on simple experimental designs is followed by regression with one dependent and one independent variable, and by factorial experiments. At this point a second pair of examinations has been customary.

Two further chapters on regression consider parallel line bioassays and associated measurements. The discontinuous Poisson and negative bionomial distributions are introduced next, leading to transformations for the analysis of variance and other ways of meeting its assumptions. I have been unable to go beyond this point in 42 lectures. The remaining topics in the syllabus have not yet been developed in as much detail as the first 17 chapters, and have varied from year to year.

More generally, the lectures consider the logic and importance of each procedure in solving illustrative biological problems, emphasizing experimental design in each case. Many of the underlying assumptions are expressed in simple mathematical models. Thus in the analysis of variance the additive model, the distinction between models I and II, and variance components are introduced at an early stage. Additivity is demonstrated numerically by isolating for each constituent a table of differences, which, when squared and summed, leads to its sum of squares in the analysis of variance. Many equations for basic statistics, such as χ^2 and the regression coefficient, are presented in several forms, suitable for computing data collected in different ways. The advantages of adapting the equation to fit each major type of problem rather than adapting the data to fit a single equation outweigh in my opinion the apparent simplicity of a single general equation.

Syllabus. The general headings and principal subdivisions of the course are summarized in the following syllabus. In outline form, the first 17 chapters vary in length of text from 3 to 13 pages.

1. Computing instructions: points in arithmetic, symbolism, number of significant figures, operation of desk calculators, use of statistical and computing tables.

2. A taste experiment: underlying concepts, design, interpretation based on the null hypothesis, criteria of rejection, relation to probability and randomization, algebra of combinations.

3. The binomial distribution: some characteristics, sample and population defined, structure of analysis, expected and observed frequencies, parameters and statistics of the binomial.

4. The χ^2 distribution: characteristics, the theoretical distribution, comparison of observed and expected frequencies, comparison of binomial statistics and parameters.

5. Analysis of proportionate frequencies: χ^2 test for $2 \times k$ con-

tingency tables, four-fold or 2×2 tables, including models, Yates' correction, Fisher's exact text, and measures of association.

6. The normal distribution: characteristics of a normal variate, relation to other distributions, theoretical normal distribution, graphic tests for normality in large and small samples, grouping, graphic estimation of mean and standard deviation, transformation to a normal metameter, contaminated and truncated distributions.

7. Numerical analysis of normal samples: criteria for adequacy of a statistic, statistics from individual observations and from a frequency distribution, non-efficient estimates, precision, comparison of observed and expected frequencies, tests of skewness and kurtosis, the rejection of outliers.

8. Interval estimation: interval estimates, Student's t-distribution, limits for the mean and for the variance, confidence vs. fiducial intervals, graphic limits for the mean.

9. The comparison of two groups: logical basis, comparison of two variances, the F distribution, comparison of two means, analysis of variance, a ranking test.

10. The comparison of several groups: structure of the comparison, tests for homogeneity of the variances, a quick test for comparing group means, analysis of variance, Models I and II for the analysis of variance, just significant difference and range, variance components.

11. Simple experimental designs: comparison of paired treatments, randomized groups or blocks, Latin squares, split plots, missing values.

12. Regression: assumptions and objectives, linear regression equations, analysis of variance of linear regressions, sampling errors, transformations to linear form, non-linear regression with orthogonal polynomials.

13. Factorial experiments: advantages, types of factor, analysis of two-factor experiments, experiments with three or more factors, error term, control of heterogeneity.

14. Bioassays from parallel regressions: role of the dosage-response curve, types of bioassay, potency from parallel log-dose response lines, factorial determination of potency, precision of the estimated potency, assays with two or more unknowns, replicated assays.

15. Associated measurements: statement of a typical problem, linear functional relations, bivariate normal distribution, correlation coefficient, significance of observed correlations, partial correlation, graphic tests for association.

16. Two discontinuous distributions: constant vs. varying expectations, Poisson distribution, χ^2 tests, indirect estimates of the Poisson parameter, negative binomial distribution, tests for agreement, estimating the negative binomial k from several series.

17. Meeting the assumptions of the analysis of variance: objectives and assumptions of the analysis of variance, non-additivity in a cross-classification, transformations for discrete variates and for continuous variates, other methods for controlling heterogeneity in the error.

18. Additional comparisons of proportionate frequencies.

19. Probit analysis for all-or-none data.

20. Covariance.

21. The comparison of slopes and slope-ratio assays.

22. Partial regression and discriminant analysis.

23. Additional designs for controlling heterogeneity.

24. Components of variance and the combining of experiments.

25. Some sampling techniques.

The several chapters in the syllabus are not of equal difficulty. In an attempt to assess this factor, my last two classes were asked in their final examinations to rank the first 17 chapters in order of increasing difficulty. The individual rankings from eight students in 1952–53 and from four students in 1954 were converted to normalized scores (4) and averaged separately for each year. The chapters and mean scores have been listed in Table 3 in order of increasing difficulty as determined from the average of the means in each set, although the class rankings differed in detail. The "just significant range" beneath each column in the table has been computed by Keul's definition (5) for comparisons of two to five items.

The individual scores for each year have been examined by the analyses of variance in Table 4. The more recent the chapter the more difficult it was judged (row 1), the trend being more pronounced in 1954 than in 1952–53. After allowing for this trend, the chapters still varied significantly from one another in difficulty for both classes (row 2). The correlation between mean scores for the two years, after removing the trend on order, was suggestive but not significant ($r = 0.41$). A comparison of the error mean squares indicates greater agreement among members of the second, smaller class.

Student comments. Four out of five questionnaires contained comments and suggestions concerning the course. About one former student in three testified as to the value of the course in his career, especially in the design of experiments, in the evaluation of data, and in his ability to read the literature critically. One of the best students placed the blame for any difficulties he had experienced upon himself. One who audited the course reported that, "I hardly see how I could operate without it." Another commented, "that your course has given me a decided professional advantage over most of my colleagues, who are for the most part abysmally ignorant regarding biometry", a viewpoint expressed by several others. One commented that, "The single

TABLE 3

Chapters 1–17 in syllabus arranged in order of increasing difficulty from average of the mean scores for two classes, with the just significant range for comparing 2 to 5 items in each year (5).

Chapter No.	Mean score 1952–53	1954	Subject
1	−1.36	−1.79	Computing instructions
2	−.87	−.99	A taste experiment
6	−.08	−.99	Normal distribution
3	−.47	−.83	Binomial distribution
7	−.46	−.56	Calculation of normal samples
9	−.75	.12	Comparison of two groups
10	−.39	.26	Comparison of several groups
4	−.08	.04	χ^2 distribution
5	.32	−.34	Contingency tables
11	−.20	.38	Simple experimental designs
8	.26	−.09	Interval estimation
15	.52	.36	Associated measurements
16	.41	.56	Distribution of counts
14	1.26	.04	Log-ratio bioassays
17	.34	.98	Meeting the assumptions of Anova
13	.71	1.67	Factorial experiments
12	1.29	1.19	Regression
2	.65	.66	Just significant range for 2 to 5 means
3	.78	.79	in each column.
4	.86	.87	
5	.92	.93	

TABLE 4

Analysis of variance of the scores averaged in Table 3.

Term	1952–53 D.F.	M.S.	F	1954 D.F.	M.S.	F
Trend on order of study	1	34.712	15.43	1	31.825	28.48
Chapters around trend	15	2.249	5.17	15	1.118	5.20
Remainder	112	.435		48	.215	

most important contribution to my own preparation was to lay before me the logical method of attacking a scientific problem, whether the experiments are to be analyzed statistically or not." No doubt every teacher of biometry has received similar reports from former students.

Many comments were more critical. Opposite changes in emphasis

have been recommended. Eight comments are typified by the following: "More fundamental theory would have helped, I got a sense of using a cook book in solving problems, which is annoying." Others expressed the opposite view. A former assistant reported "biologists didn't seem too keen about theoretical matters, they wondered why and how an F value was reached but didn't like to go into the mathematics of it". One comment reads, "the most important aspect of biometry from my experience is proper application, how the formulae are derived is not as important in this applied field".

Several wanted more specialization within branches of science, but others were willing to settle for more examples from which to choose. Only four comments questioned rather mildly the desirability of teaching a single course in biometry to biologists from quite different fields, a basic feature of the course. In at least one case, taking the course changed the student's original viewpoint as indicated by the following comment: "Although I disagreed with you on this point, I want to congratulate you on having the courage to conduct a course for all biologists. I disagree completely with those who 'pigeonhole' the different fields of biology." In another reply, this basic tenet is considered to present a major difficulty "in the fact that each student brings such a widely different background to the course. While the skeletal outline of the course is adequate, each student must be given special attention to a degree not warranted in other courses and this attention must be a function of his background, needs, tastes and objectives. Perhaps an impossible order but no other course has this inherent and unfortunate complexity."

Some suggestions concerned the laboratory. Two wanted examples of the misuse of statistical technique and instruction in what *not* to do as well as in what to do. One wanted more emphasis upon evaluating the method used in obtaining biological data. One proposed "that the laboratory exercises include in each section at least one very simple example, one in which the arithmetic is so very simple that the calculations can be followed without a calculator and at home." A former assistant spent "considerable time teaching the students elementary statistics before they could proceed to the assigned examples. As long as provision is made for this supplemental teaching, I would recommend continuing a highly concentrated course instead of diluting it." How the laboratory assistant should spend his time is debatable. He might discuss informally with small groups of students the mathematical background of biometrical relations to compensate for their gaps in basic mathematics. Alternatively, he could concentrate on practical advice on actual computations and on the biological interpretation of results. The latter is the primary intent of laboratory instruction,

although only the laboratory assistant may be able to appreciate which background concepts need explaining to a particular student.

The most frequent criticism from a dozen or more former students was that "too much material was covered in too short a period of time." Another wrote "one had the feeling he was just coming out of a fog when the instructor rushed onto something new and lost us again". The remedy suggested most frequently was "to extend the course to a full year rather than a semester" and this recommendation has now been adopted with two one-hour lectures a week.

One other concept appeared in several questionnaires. This is the idea of a delayed response in learning statistics. One former student wrote, "I entered this course completely ignorant of statistical techniques and theory. I was somewhat confused and bewildered when I finished. However, since returning to my normal work, I find that I can work fairly well in a statistical sense in my own field and more than hold my own statistically with my associates." A second commented: "What stands out in my mind, even though it is now six years later, is the extreme practicability of the course. I am astonished now that I got as much as I did, since my feeling when attending the course was often one of 'can't see the woods for the trees'." Another wrote: "I suspect that I never began to feel 'easy' about even the simplest statistical methods until I began to try to teach a little statistics to medical students." One who is now teaching a semester of biometry to zoology majors is "convinced that during the course you can only *expose*. The real learning comes later through solving particular problems. Only later does the usefulness of the course become evident. At the end of your course I would have rated it as mediocre, but ever since I have had a clarity of thinking in the field that is extremely useful." These last comments suggest that in teaching biometry there is a latent period between the stimulus of teaching and the response of learning and that the real effectiveness of a course cannot be judged until later.

REFERENCES

1. Bliss, C. I. and Calhoun, D. W., *An Outline of Biometry*, Yale Co-Operative Corporation, New Haven, 1954.
2. Fisher, R. A., *Statistical Methods for Research Workers*, Eleventh Edition, Oliver and Boyd, Edinburgh, 1950.
3. Fisher, R. A., *The Design of Experiments*, Sixth Edition, Oliver and Boyd, Edinburgh, 1951.
4. Fisher, R. A. and Yates, F., *Statistical Tables for Biological, Agricultural and Medical Research*, Fourth Edition, Oliver and Boyd, Edinburgh, 1953.
5. Keuls, M., *Euphytica*, 1: 112, 1952.
6. Walker, H. M., *Mathematics Essential for Elementary Statistics*, Revised Edition, Henry Holt and Company, New York, 1951.

THE THEORY OF BACTERIAL CONSTANT GROWTH APPARATUS

C. C. Spicer

Central Public Health Laboratory, Colindale, London, N.W.9

Recent work in bacterial genetics has emphasised the usefulness of apparatus whose purpose is to maintain a constant population of bacteria in a state of active growth. Several such devices have been described for example by Monod (1950), Novick and Szilard (1950), and Perret (1954).

It seems worthwhile to give a short account of the underlying theory as a guide to the problems of design likely to be encountered with organisms of different growth characteristics; or under various conditions of culture.

Mathematically the problem may be stated as follows:—Consider an organism growing freely in a limited, constant volume of nutrient. After some period of growth factors come into play which depress the power of the organism, to divide and eventually stop it growing altogether. These factors may be of several different kinds: for example, exhaustion of nutrient, insufficiency of oxygen, or production of some toxic metabolite, the general effect however, is that the growth rate of the population at any time is a function of its size (n) so that

$$\frac{1}{n}\frac{dn}{dt} = f(n)$$

where $f(n)$ is some function of n. If the organism is growing in some apparatus which is constantly renewing the medium and concurrently removing a fraction β of the organisms per unit time, the equation of growth becomes:

$$\frac{1}{n}\frac{dn}{dt} = f(n) - \beta$$

In a constant growth apparatus

$$\frac{1}{n}\frac{dn}{dt} = f(n) - \beta = 0$$

and the washing-out rate required to maintain a population of a given size is found by solving the equation $f(n) = \beta$. Such an equilibrium is not necessarily stable. For instance, if the population is growing exponentially it is not possible in practice to maintain a constant number by simply renewing the medium, as small discrepancies between

225

the growth rate of the organism and the turnover of the medium always occur which result in either washing out of the organisms or over growth.

In general, if the equilibrium population $\tilde{n}$ by some accident becomes $(\tilde{n} + \eta)$, where η is small compared with $\tilde{n}$ we have

$$\frac{d(\tilde{n} + \eta)}{dt} = (\tilde{n} + \eta)f(\tilde{n} + \eta) - \beta(\tilde{n} + \eta)$$

If η is small $f(\tilde{n} + \eta)$ can be expanded by Taylors Theorem and, ignoring terms in η^2 and higher powers of η,

$$\frac{d\eta}{dt} = (\tilde{n} + \eta)(f(\tilde{n}) + \eta f'(\tilde{n})) - \beta(\tilde{n} + \eta)$$

At equilibrium $\beta = f(\tilde{n})$

$$\frac{d\eta}{dt} = \eta\tilde{n}f'(\tilde{n})$$

Now if the equilibrium is to be stable any change in η must cause an opposite change in $d\eta/dt$, i.e. if η is positive $d\eta/dt$ must be negative and vice versa. So, in general, equilibrium is only stable if $f'(\tilde{n})$ is negative. In other words there can be no stability unless the growth rate decreases as the concentration of organisms increases.

The most completely worked out system so far used is the chemostat of Novick and Szilard (1950a, 1950b). This applies to an organism dependent on a nutrient factor present in such a limiting quantity that small variations in concentration can cause corresponding variations in growth rate. Then, if c is the concentration in the growth tube, the equation for the growth of the organism is

$$\frac{1}{n}\frac{dn}{dt} = F_1(c) - \beta$$

and the corresponding equation for changes of concentration is

$$\frac{dc}{dt} = \beta(a - c) - F_2(n, c)$$

Here, a is the concentration of nutrient in the incoming medium, and $F_2(n, c)$ is a function describing the rate at which the nutrient is taken up by the organism.

Novick and Szilard have shown, for several nutrient factors, that over a certain range of c we can write

$$F_1(c) = \lambda c$$

$$F_2(n_1 c) = \kappa nc$$

Where λ and κ are constants. A similar approximation should hold for any nutrient in limiting concentration.

The differential equations of growth under these circumstances are

$$\frac{1}{n}\frac{dn}{dt} = \lambda c - \beta$$

$$\frac{dc}{dt} = \beta(a - c) - \kappa nc$$

At equilibrium the concentrations of organisms ($\tilde{n}$) and nutrient ($\tilde{c}$) can be found by equating the derivatives to zero, which gives

$$\tilde{c} = \frac{\beta}{\lambda}$$

$$\tilde{n} = \lambda(a - \tilde{c})/\kappa$$

The general solution of the two simultaneous non-linear differential equations for n and c cannot be conveniently given in general terms. It is possible however, to investigate the response to small displacements from the equilibrium position. In the region of equilibrium put $n = (\tilde{n} + \eta)$, $c = (\tilde{c} + \xi)$ where η and ξ are so small that their squares and product may be neglected. Substituting these variables in the growth equations it is found that

$$\frac{d\eta}{dt} = \lambda\tilde{n}\xi$$

$$\frac{d\xi}{dt} = -(\beta + \kappa\tilde{n})\xi - \kappa\tilde{c}\eta$$

The solution of this pair of equations can be put in the form

$$\eta = A_1 e^{\mu_1 t} + A_2 e^{\mu_2 t}$$

$$\xi = B_1 e^{\mu_1 t} + B_2 e^{\mu_2 t}$$

where the coefficients A and B are determined by the initial conditions and μ_1, and μ_2 are the roots of the quadratic equation

$$x^2 + \lambda ax + \lambda\beta(a - \tilde{c}) = 0$$

Now, so long as $a > \tilde{c}$, which is necessary if a constant population is to be maintained, the roots of this equation are real and negative. Consequently any small displacement from equilibrium dies away exponentially without oscillations and the steady state is always stable.

It is worth pointing out that if a population of organisms grows according to the conditions specified by Novick and Szilard, but without washing out or removal of nutrient then its growth curve follows the

well known "Logistic Law". Eliminating c from the differential equations of growth we have

$$\frac{1}{n}\frac{dn}{dt} = (\lambda c_0 + \kappa n_0) - \kappa n$$

where n_0 and c_0 are the initial concentrations of organism and nutrient. This is the equation of a logistic population whose final size, is

$$n_\infty = n_0 + \frac{\lambda c_0}{\kappa}$$

The number of organisms which have been produced from a unit concentration of nutrient is then

$$\frac{n_\infty - n_0}{c_0} = \frac{\lambda}{\kappa}$$

So that λ/κ is the amount of growth factor required to make a single organism.

Writing the equation of the logistic curve in the form

$$N = \frac{N_\infty}{1 + e^{-\epsilon t}}$$

with the origin of t at the time when $N = N_\infty/2$, then the value of the constant ϵ is given by

$$\epsilon = \lambda c_0 + \kappa n_0$$

As a first approximation to the behaviour of the organism in a constant growth apparatus, it can be considered to be growing logistically but being at the same time washed out, so that

$$\frac{1}{N}\frac{dN}{dt} = \epsilon\left(1 - \frac{N}{N_\infty}\right) - \beta$$

Under the general stability conditions $f'(n)$ is negative and equilibrium is always stable, also

$$\tilde{n} = \left(1 - \frac{\beta}{\epsilon}\right)n_\infty$$

so that theoretically any desired population $< n_\infty$ can be maintained. However, the smaller the population the greater β must be, and unless it is regulated with great accuracy it is liable to exceed ϵ and the population will then be washed out.

It is probable that the equilibria attained with rather poor media will all be of the general type discussed here. The general form of the dependence of growth rate on a limiting nutrient is approximately

exponential so that

$$F_1(c) = \lambda(1 - e^{-kc})$$

As the concentration of growth factor is increased it ceases to have an increasing effect on growth, while at low concentrations its effect is approximately linear. Differential equations of growth which contain this form for $F_1(c)$ have a stable equilibrium similar to that for the simple linear form, but with a different time constant. There are no oscillations about the equilibrium.

As a contrast to the case of growth limited by shortage of growth factor discussed above it is worth considering a simple model of a population which is limited by production of some toxic substance. There is no example of this kind that has been so well worked out as Novick and Szilard's nutrient scheme, but there is no doubt that toxic limitation can occur. It could be imitated in a constant growth apparatus by adding an antibiotic at a rate governed by the density of organisms.

Taking only the simplest case, the differential equations of the system would be

$$\frac{1}{n}\frac{dn}{dt} = (\lambda - \beta) - \mu c$$

$$\frac{dc}{dt} = \gamma n - \beta c$$

The constants μ and γ here represent the lethal effect of toxin on the organisms and its rate of production by them. In the absence of washing-out ($\beta = 0$) the organisms eventually become extinct while the concentration of toxin rises logistically to a constant value. Under constant growth conditions an equilibrium is established when

$$\tilde{c} = \frac{\lambda - \beta}{\mu}$$

$$\tilde{n} = \frac{\beta\tilde{c}}{\gamma}$$

The equations for determining the stability of the equilibrium are

$$\frac{d\eta}{dt} = -\mu\tilde{n}\xi$$

$$\frac{d\xi}{dt} = -\beta\xi + \gamma\eta$$

where, as before, η refers to the disturbance of bacterial numbers and ξ to that of toxin concentration. The solutions are of the exponential form given above, and the coefficients of t in the exponents are the

roots of the quadratic equation

$$x^2 + \beta x + \mu \tilde{n} \gamma = 0$$

When $\beta > 4/5 \; \lambda$ both roots are negative and real and the equilibrium is stable. If $\beta < 4/5 \; \lambda$, then the roots are imaginary with negative real parts and the equilibrium is still stable, but is reached by damped oscillations.

The general equations of equilibrium, which are applicable to both toxic or nutritional schemes, can be derived from the two differential equations.

$$\frac{1}{n} \frac{dn}{dt} = F_1(c) - \beta$$

$$\frac{dc}{dt} = \beta(a - c) + F_2(c, n)$$

by expanding the functions F_1 and F_2 about the point $(\tilde{n}_1 \tilde{c})$ in a Taylors series. This procedure gives, for small displacements

$$\frac{d\eta}{dt} = \xi \tilde{n} \frac{\partial F_1}{\partial c}$$

$$\frac{d\xi}{dt} = \xi \left(\frac{\partial F_2}{\partial c} - \beta \right) + \eta \frac{\partial F_2}{\partial n}$$

The quadratic equation whose roots are the coefficients of t in the solution is

$$x^2 - \left(\frac{\partial F_2}{\partial c} - \beta \right)x - \tilde{n} \frac{\partial F_1}{\partial c} \cdot \frac{\partial F_2}{\partial n}$$

For stable equilibrium $\partial F_1/\partial c$ and $\partial F_2/\partial n$ must be of opposite sign and if $\partial F_2/\partial c > 0$ then $\partial F_2/\partial c < \beta$.

Summary. A mathematical analysis is presented of the mechanism of certain types of bacterial constant growth apparatus. The conditions of equilibrium and the nature of response to displacements from it are derived.

Acknowledgements

References

Monod, J. La technique de culture continue. *Ann. Inst. Pasteur 79*: 390 (Oct.) 1950.

Novick, A., and Szilard, L., (a) Experiments with the Chemostat on spontaneous mutations of bacteria, *Proc. Nat. Acad. Sci. 36*: 708.

Novick, A., and Szilard, L., (b) Description of the Chemostat, *Science 112*: 715.

Perret, J., (1954). In the press.

AN INVERTED MATRIX APPROACH FOR DETERMINING CROP-WEATHER REGRESSION EQUATIONS*

Harold F. Huddleston

U.S.D.A. Agricultural Marketing Service
Washington, D. C.

Introduction

We would like to know whether year-to-year changes, or month-to-month changes in crop yields or prospects are consistent with observed weather data. Generally, historical weather records extend back farther than records of crop yields. We wish to make use of weather data for the entire period of record even though yield data may be available for a much shorter period. This paper reports on an exploratory inverted matrix approach used in one phase of a crop-weather study.

The application of multiple regression methods in the study of relationships between crop yields and weather factors is, of course, not new, but the large amount of computational labor involved has discouraged many workers and our people from attempting correlations studies on a very extensive scale. As pointed out by R. A. Fisher, the use of the inverse matrix solution of a set of normal equations greatly reduces the amount of computations when the same set of independent variables is used repeatedly; in addition, it serves to simplify the calculation of sampling errors of the regression coefficients. However, a large amount of computational work is still required when the various dependent variables are available for only relatively few years, and these periods vary from crop to crop because of the fact that the data or series were started at different points in time. We would like some way of utilizing all the weather and crop yield data available. Therefore, we would like to devise what might be called "generalized inverse matrix solution" for a given State or area which could be used whenever the given set of weather factors were appropriate. However, the sampling errors of the regression coefficients cannot be computed using the elements of this generalized solution where the dependent variables are used for only a subperiod.

The inverse matrix solution is obtained for a given set of independent variables (i.e., weather factors) for the entire period of the weather

---

records, or at least some fairly long period of years. It is found that
the elements of the inverse matrix exhibit stability as the length of the
period is increased. The elements or their ratios are used with the
covariance terms between yields and each of the weather factors for
the various subperiods for which crop yield data are available. Obvious-
ly, the underlying assumption which is made is that the interrelationship
between given weather factors will remain fairly constant and become
more reliable over time. In the example used it is assumed that this
stability is over years for fixed months.

Nature of Study

In order to clarify the ideas and procedures suggested in the present
study, an example of an application is given. The State of Illinois has
arbitrarily been selected for examination and illustrated for corn yields.
The study was conducted in the following manner. A linear relation-
ship between yields and monthly rainfall and temperature data was
used. Linear regressions have been found to give fairly satisfactory
results in many cases for these variables. The functional relationship
used was as follows:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

Where

Y = yield per acre

X_1 = average monthly rainfall for State.

X_2 = average monthly temperature for State.

X_3 = product of average monthly rainfall and temperature for the
State, or $X_3 = X_1 \cdot X_2$ neglecting decimals (i.e., the product
of rainfall and temperature).

Since in many multiple regression studies, joint effects may be
important, the product (X_3) of rainfall and temperature was included
as a third factor. The utility of the third factor has been pointed out
by Hendricks[1] and Scholl where an understanding of the effects of
weather is of interest and its inclusion appears desirable for a generalized
regression approach.

The period selected for study was 1891–1950. The rainfall and
temperature data for the month of July were selected for examination.
The inverse matrix solutions were computed for the following sub-
periods as well as the entire period: (1) 1891–1910; (2) 1911–1930; (3)
1931–1950; and (4) 1911–1950. Table I indicates the C_{ij} values for

---

[1] Agricultural Experiment Station Technical Bulletin # 74 "Techniques in Measuring Joint
Relationships" by North Carolina State College 1943.

TABLE I

Inverse Solutions—Illinois July Weather Data

C_{ij}	20 Year Periods			40 Year Period	60 Year Period
	1891–1910	1911–1930	1931–1950	1911–1950	1891–1950
C_{11}	+3.4509	+1.2288	+7.7944	+14.001	+5.0598
C_{12}	+0.14000	+0.093251	+0.23438	+ 0.45139	+0.16964
C_{13}	−0.047482	−0.015188	−0.10123	− 0.18371	−0.066759
C_{22}	+0.015796	+0.022006	+0.014153	+ 0.019131	+0.0087353
C_{23}	−0.0017839	−0.00091567	−0.0030412	− 0.0058850	−0.0022128
C_{33}	+0.00062649	+0.00019709	+0.0013209	+ 0.0024136	+0.00088277

the various periods, where the C_{ij} are defined by the following set (i.e., $j = 1, 2, 3$) of equations where $k = 20, 40, 60$:

$$\sum_{}^{n_k} x_1^2 C_{1j} + \sum_{}^{n_k} x_1 x_2 C_{2j} + \sum_{}^{n_k} x_1 x_3 C_{3j} = 1, 0, 0$$

$$\sum_{}^{n_k} x_1 x_2 C_{1j} + \sum_{}^{n_k} x_2^2 C_{2j} + \sum_{}^{n_k} x_2 x_3 C_{3j} = 0, 1, 0$$

$$\sum_{}^{n_k} x_1 x_3 C_{1j} + \sum_{}^{n_k} x_2 x_3 C_{2j} + \sum_{}^{n_t} x_3^2 C_{3j} = 0, 0, 1$$

A study of the values in Table I indicates the absolute values of the C_{ij} will vary considerably from one period to the next. However, the ratios of the C_{ij} to each other are of interest in studying the tendency for stability of interrelationships of weather factors. In addition, Table II below shows the ratios of C_{ij} to C_{11} for each period.

TABLE II

Ratios C_{ij} to C_{11}

C_{ij}'	20 Year Periods			40 Year Period	60 Year Period
	1891–1910	1911–1930	1931–1950	1911–1950	1891–1950
C_{11}'	1.00000	1.00000	1.00000	1.00000	1.00000
C_{12}'	.04056	.07589	.03007	.03224	.03353
C_{13}'	− .01376	− .01236	− .01299	− .01312	− .01319
C_{22}'	.004577	.01791	.001816	.001366	.001726
C_{23}'	− .0005169	− .0007452	− .0003902	− .0004203	− .0004373
C_{33}'	.0001815	.0001604	.0001695	.0001724	.0001745

An inspection of the ratios reveals several things; (1) Relationships based upon 20 years of weather data may be expected to have little reliability for subsequent years; (2) in general, it would seem advisable that relationships using weather data for as large an area as a State should be based upon at least 40 years of data in order to obtain stable relationships among the weather factors; and (3) the ratios are fairly stable from period to period in contrast to their absolute values.

The utility of the multipliers for a long period of years which could be used as "population values," i.e., C'_{ij}, as indicated by this analysis appears to be dependent upon: (1) Finding a quick method of estimating a factor of proportionality, K, by which one can convert the ratios to absolute units, or (2) using the ratios of the C_{ij}, as in Table II, to compute regression coefficients proportional to the net regression coefficients; then obtain the relationship between yields and the weather factors by plotting the computed regression values (using the proportional regression coefficients) against the actual yields or deviations from the average yield. Further study of the variances and covariances involved appears necessary before any conclusion can be made concerning the feasibility of determining a suitable value of K a priori.

The multipliers in Table I for any of the periods may be used with any number of crop yields for the same period for the State by computing the respective covariance terms. The computational work is, therefore, considerably reduced. The data in Table II for the 60-year period (last column on right) is thought of as a "general solution".

As an example of the use indicated in (2), the yield of corn is correlated with the July weather data for Illinois. The proportional net regression coefficients are computed as follows:

$$b'_{y2.34} = C'_{11} \sum_{}^{n} x_1 y + C'_{12} \sum_{}^{n} x_2 y + C'_{13} \sum_{}^{n} x_3 y$$

$$b'_{y3.24} = C'_{12} \sum_{}^{n} x_1 y + C'_{22} \sum_{}^{n} x_2 y + C'_{23} \sum_{}^{n} x_3 y$$

$$b'_{y4.23} = C'_{13} \sum_{}^{n} x_1 y + C'_{23} \sum_{}^{n} x_2 y + C'_{33} \sum_{}^{n} x_3 y$$

Where $\sum_{}^{n} x_1 y$, $\sum_{}^{n} x_2 y$, and $\sum_{}^{n} x_3 y$ are sums of products of deviation from means for the yield of corn per harvested acre (Y) with the monthly averages of rainfall (X_1), temperature (X_2) and the product of temperature and rainfall (X_3) for the period 1911–1950 after the yields have been adjusted for trend (i.e., by use of 10-year averages). The C'_{ij} used are based upon the period 1891–1950 in equation 1) and 1911–1950 in equation 2). The regression equation for computing values from the proportional regression coefficients is:

$$Y'_c = b'_{y1.23} x_1 + b'_{y2.13} x_2 + b'_{y3.12} x_3$$

or

1) $Y'_c = -2.315x_1 - .1295x_2 + .0375x_3$

The actual regression between the adjusted yields and weather factors determined from the data for the period 1911–1950 is given by the following equation:

$$Y_c = b_{y1.23}x_1 + b_{y2.13}x_2 + b_{y3.12}x_3$$

or

2) $Y_c = -24.59x_1 - 1.214x_2 + .353x_3$

The values of Y_c and Y'_c are plotted against $Y - Y_T$ (deviations from trend) in Chart I.

An inspection of Chart I indicates that there is little difference in the relationship found by use of the actual data for the period 1911–1950 and the ratios of the C_{ij} for the period 1891–1950 with the covariance terms for 1911–1950.

A factor (K) which can be used to covert the proportional regression coefficients to the actual values will be equal to the slope of the regression line in Part B of Chart I. That is, K (determined by least squares method) multiplied by the proportional regression coefficients will give the regression in absolute units. A comparison of the coefficients in equations 1) and 2) indicates a factor of about 10 is needed to convert the proportional coefficients to an absolute basis.

The C_{ij}'s (or C'_{ij}'s) in column 5 of Table I (or II) can similarly be used with various subperiods corresponding to the years for which the individual crop yield data are available. However, we would prefer, in general, to express the yield data as a percent of the normal or average yield rather than as deviation in absolute units. If the yield data are expressed in the percentage form, year-to-year changes are indicated by the ratio of the two years. The percentage change can then be converted to bushels per acre rather than determining the regression coefficient in their true or absolute units.

Conclusions: While a fairly large amount of computational work is involved in any multiple regression technique, it is believed that a generalized regression approach may be useful in many situations. The utilization of lengthy weather records to establish stable relationships among weather factors with determination of the covariance terms where yields are available for a much shorter period of time would appear practical based on preliminary results. In addition, the time and costs required to compute the inverse matrix solutions are not nearly so formidable with the aid of modern computing machines as has been the case in the past. It is possible that if work can be expanded

CHART I RELATIONSHIP OF ILLINOIS CORN YIELDS WITH WEATHER FACTORS - 1911-1950
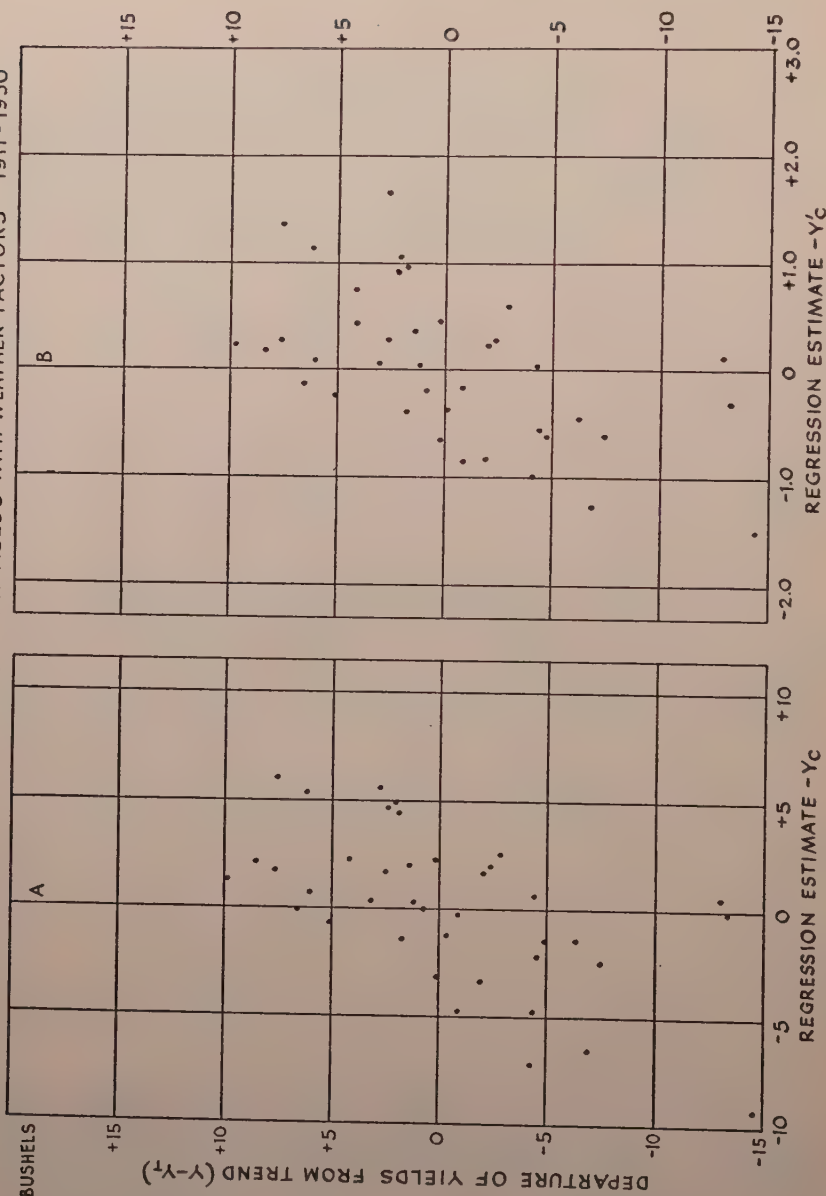
along these lines a more objective means for estimating the effects of weather factors on crop production from available weather records can be used to supplement the current procedures of the Crop Reporting Board.

QUERIES

QUERY: A recent query (107, March, 1954) presented an inter-
114 esting discussion of some points on Sheppard's correction. I
would like to raise some additional points on application of the
correction in making tests of differences between means or analysis of
variance tests. The pertinent reference again is Fisher. I also checked
M. G. Kendall's "Advanced Theory of Statistics".

In my case I was supplied with a set of data in frequency distribution
form. Unfortunately, the class interval was rather wide, 200 units,
while the estimated standard deviation was about 270 units (based on
the grouped data). On the other hand, the data included the means
calculated from the original ungrouped observations for each treatment
combination.

After completing the analysis without correction, it occurred to me
that perhaps the matter of Sheppard's correction should be considered.
Hence, I checked the references noted above, but was not satisfied
with the information obtained. That is, I was not told exactly why the
correction was not to be applied for tests of significance even though it
seemed to be appropriate for estimation.

In my situation it appeared to me that since I had means based on
original data it might be appropriate to apply the correction for esti-
mating the variance of a difference between means. Upon carrying
out the necessary calculations, I found the correction to the second
moment to be large, but the actual effect on the final value of Student's
t or a normal deviate, Z, to be negligible.

In discussing the matter with a colleague this point of view was
suggested: When both the mean and standard deviation are calculated
from a grouped frequency distribution, the two statistics are both in
error by some amount and the direction of the error for the mean is
unknown. Thus, one might recommend, as does Fisher, "do not apply
the correction for tests of significance" and the long-run results should
be all right.

Question: (1) What is the real basis for Fisher's advice? and (2)
Was I right in not applying Sheppard's correction for my case?

ANSWER: The basis of Fisher's advice was that grouping introduces
an additional component of variance of which the magni-
tude is known on the assumption of perfect grouping, e.g.
that the true measurements of those classed as 17 units do all lie between
16.5 and 17.5 exactly, and are all that lie between those limits. For an

analysis of variance the effect is simply to add this fixed quantity to all mean squares, so reducing the probability that they should be unequal at any chosen ratio. In effect, errors of grouping, like other errors of random sampling, lower the precision with which any comparison can be made. Their exact and particular effects are always unknown, although the average magnitude is known, and is what is removed from the variance in making Sheppard's correction.

In your case errors of grouping have not been introduced in calculating the means to be compared, but only in calculating the estimate of error. I should, in such a case, apply the correction to the latter before testing the significance of the former.

R. A. FISHER

QUERY: Marvin Zelen in a recent issue of Biometrics (p. 273, **115** Vol. 10) states "almost every experiment in the physical sciences is characterized by the block being a 'natural experimental unit'." This terminology is not in accordance with the generally accepted (?) idea that the experimental unit is part of the array of experimental material (including perhaps a classification of material by time or other extraneous attributes) wich receives a treatment independently of other parts within the restrictions of the design? What exactly does the term "natural experimental unit" mean?

ANSWER: I am not quite certain that I fully understand the query. However I shall amplify my statement concerning "natural experimental units" in the hope that this will also satisfactorily answer the query. First to quote Cochran and Cox, in their book *Experimental Designs* (p. 15), "We shall use the term *experimental unit* to denote the group of material to which a treatment is applied in a single trial of the experiment. The unit may be a plot of land, a patient in a hospital, or a lump of dough, or it may be a group of pigs in a pen, or a batch of seed."

The reason for using the adjective *natural* was to further emphasize that in physical science applications, the block arises because of some natural grouping of the experimental material or because of limitations in applying the different treatments. On the other hand, in many agricultural field trials a plot of land is selected for the experiment and the land is arbitrarily partitioned into blocks for the purpose of the experiment. That is, the partitioning of the land into blocks or units is not unique and usually depends on the convenience of the individual who is planning the work.

In many experiments there is a natural limitation within the experiment itself which determines the block. For example, in an experiment on eye preparations which are to be tested on humans, the block might consist of an individual and thus only two different preparations could be applied within any one block, one for each eye. The eye is the experimental unit, they come in pairs to form a "natural" block, and there is nothing the experimenter can do to change the situation, unless of course, he has access to three-eyed people.

MARVIN ZELEN

QUERY: W. T. Federer and C. S. Schlottfeldt in a recent issue
116 of Biometrics (Vol. 10, p. 290) state, "The decision to use covariance to control gradients after the experimental results have been studied invalidates the use of tabulated probability values for the standard tests of significance". Can the authors elaborate this statement further? The statement appears to contradict some of R. A. Fisher's writing; e.g. in *Design of Experiments* and much that is written in texts on statistical methods.

ANSWER: The statement referred to above does not contradict the material in R. A. Fisher's *The Design of Experiments* or in statistics texts. To illustrate consider that two six-sided dice are to be cast singly. Now, if one die is observed first before placing bets then the cast of the second die is all that is important. For example, suppose that a six is observed on the first die. Now, the probability of obtaining any number between 7 and 12 on the two dice is 1/6. That is, only the result of the second die counts in computing the probabilities. The probabilities of obtaining the numbers 2 to 12 resulting from casting two dice simultaneously cannot validly be used for the "result guided procedure" described above.

If the experimental results are studied to determine which covariate will reduce the experimental error, the tabulated probability levels for t, z, F, χ^2, etc. cannot validly be used to test these experimental results. However, if the experimenter decides on the covariate *prior* to studying the experimental results, then the tabulated levels of the various tests of significance may validly be used.

Little published material is available on the problem of first studying the experimental results and then deciding what to do next. Dr. T. A. Bancroft, Iowa State College, and his students have made a start on the problem of using "result guided procedures".

W. T. FEDERER

ABSTRACTS

304 SAMUEL W. GREENHOUSE. (National Institute of Mental Health and George Washington University.) **Information and Distance Applied to Discriminant Analysis Between Two Normal Populations.**

Given two k-variate normal populations π_1, and π_2, with parameters $\mu_{(p)}$ and $\sigma_{(p)}$ $(p = 1, 2)$, where μ is a vector of means and σ the matrix of variances and covariances, Kullback defined the mean information in an observation $X(= x_1, x_2, \cdots, x_k)$ drawn from π_1, in discriminating between π_1, and π_2 as $I(1 : 2) = \int f_1 \log (f_1/f_2)dx$. With a similar definition for $I(2 : 1)$, he defined distance as $J(1, 2) = I(1 : 2) + I(2 : 1) = \int (f_1 - f_2) \log (f_1/f_2)d_x$.

In discriminant analysis, one seeks a linear function of the x's to distinguish between π_1, and π_2. In this paper both information and distance are maximized in two situations: $\sigma_{(1)} = \sigma_{(2)}$ and $\sigma_{(1)} \neq \sigma_{(2)}$. In the former situation the same linear discriminant is obtained as that found by Fisher and is equivalent to the likelihood ratio solution. In the latter case, the same principle of maximizing information and distance is used to obtain a linear discriminant. Here, however, max $I(1 : 2)$, max $I(2 : 1)$ and max $J(1, 2)$ yield different functions. Errors of classification are investigated for each function and compared with the errors associated with linear functions obtained by other means.

305 JOHANNES IPSEN. (Harvard School of Public Health.) **Appropriate Scores in Bio-assays using Death-Times and Survivor Symptoms.**

Many bio-assays can be arranged in a $(m \times k)$ contingency table with k doses and in m categories of biological observation, ranked in order of increasing effect of treatment (e.g., survival times $\rightarrow$, survival with symptoms $\rightarrow$, and survival without symptoms). The author obtained a set of m scores that satisfies one criterion of an efficient bio-assay:

The variance of the linear regression of the mean scores on log dose is the highest possible fraction of the total variance.

A procedure is described for combining data from separate experi-

ments of similar kind, to obtain a score system so that the common slope is the maximum possible fraction of the total variance adjusted for individual means.

Significance tests for different score systems are described and the method is applied to an inter-institutional assay of tetanus toxoid comprising 96 experiments.

306 D. G. HORVITZ, J. FLEISHER, and A. L. FINKNER, (North Carolina State College.) **A Comparison of Random and Non-Random Plot Selection.**

The Agricultural Marketing Service of the United States Department of Agriculture is engaged in an extensive research program of objective sampling and measurement methods with a view toward improvement of crop acreage estimates and production forecasts. Included in the program is an investigation of the association of observable cotton plant characteristics during the growing season with final yield in order to develop a reliable production forecasting procedure. The plant data, including boll counts, are collected from small plots within sample fields.

Chain measurements of dimensions on a sample of 60 cotton fields in three North Carolina counties permitted random selection of plots within these fields and hence an evaluation of less costly non-random methods of locating similar sized plots. Four non-random plot selection schemes were examined, each scheme yielding a pair of double row plots 10 feet in length. The first of these schemes selected a border plot and an interior plot, the second selected an end of row plot and an interior plot, the third and fourth both selected a pair of interior plots. One of the four schemes was assigned at random to each sample field; two random plots were also selected from each sample field.

In addition to comparison of the mean boll counts on September 1 and at harvest, the data were analyzed to determine the contributions of the various error components to the total error. The schemes using pairs of interior plots yielded positively biased boll counts on both occasions while those consisting of a border or end of row plot and an interior plot were negatively biased. The latter schemes exhibited two to three times the variability of the schemes consisting of two interior plots. The greatest portion of this difference is accounted for by the large variability of the individual field biases for the non-random schemes using a border or end of row plot. The covariance between the individual field biases and the true field average also contributed considerably to the magnitude of the mean square errors.

The four non-random plot selection schemes taken as a group indicate undue emphasis on border and end-row plots. The statistical efficiency of the group relative to random plot selection was estimated to be 70.6% for September 1 boll counts and 61.2% for final boll counts. A distribution of the non-random plots which increases the ratio of interior plots to border and end of row plots should reduce the net bias and raise the efficiency.

M. C. K. TWEEDIE. (Virginia Polytechnic Institute.)
307 **Some Applications of a Special Lemma on Characteristic Functions.**

R. A. Fisher (Prof. Royal Soc. London, A, 144 (1934) showed that in some families of distribution functions a parameter appeared in such a position that the characteristic function could be evaluated without integration. This note applies this idea to further problems, and shows that precisely chi-square distributions can arise in more general cases than directly from normal or other chi-square distributions.

G. S. WATSON. (Australian National University, Canberra.)
308 **Contingency Tables with Missing or Mixed-Up Cell Entries.** (By Title)

In analysis of variance, missing or mixed-up entries may be dealt with by well-known methods. The same problem seems to have been overlooked in the analysis of frequency data. It is shown in this paper, however, that the method of maximum likelihood leads to easy solutions of these problems in the analysis of contingency tables.

German Section of the Biometric Society at Bad Nauheim
(Kerckhoff-Institute) January 28–30, 1955

309 R. K. BAUER, Munich. **Experiences with discriminant functions.**

Since it has been proved that the Fisher-Welch analysis yields optimum separation, the question has been settled which method of statistical balance should be used in diagnosing paternity. Ludwig suggested the application of the Penrose-Smith analysis. Then the assumptions may be weakened which have to be made on the separating traits, i.e. on the hereditability of the morphologic and physiologic items. A certain degree of freedom is gained in defining hereditability.

The most serious of the remaining assumptions, the homogeneity of variances in the collectives which have to be separated, may be approached empirically. Under unfavorable conditions it may be either enforced with usual methods or even avoided. Statistical procedures are available for the choice of the traits which used to be made authoritatively. The significance of statements may be tested, also a comparison of discriminant functions, and the size of an interval of indifference. In diagnosing paternity by using a statistical balance it becomes possible for the first time to base the plausibility of a judgement on probability theory. Special observations may be dealt with by introducing a priori probabilities.

310 H. DRUCKREY, Freiburg i.Br. **Theoretical interpretation of the processes underlying pharmacological effects.**

The relationship between dosage and effect is developed by using 'dimensional equations' in order to indicate the dimensions and mutual connections of variables on which pharmacological effects depend. At the same time an attempt is made to define basic concepts of pharmacology more precisely, e.g. poison, dosage, effect.

According to the 'theory of hits' the primary assumption is, that molecules of a poison act on particular 'receptors' of cells. The formulation of this phenomenon by using a dimensional equation corresponds in principle to the scheme for the kinetics of a bimolecular reaction. For the case of equilibrium an algebraic development yields results identical to formulae of the law of mass action, of isothermal adsorption, of diffusion, to empiric equations by A. J. Clark or A. Rosenblueth for the dosage-effect relationship, and finally to the 'logit' representation. The curves are hyperbolas. A linear function results if logarithms are taken of both members of the basic equation. A new probability grid for the dosage-effect relationship is based on this fact. At the same time it is explained that symmetric or linear functions are usually not found but by plotting versus the logarithm of dosage.

A further numerical elaboration of dimensional equations gives significant information on the dimensions of variables on which the effect depends. Even the 'individual variation' may be referred to certain variables. The effect of a poison does not depend exclusively on the dosage, but on its ratio to the number of particular receptors in the effective volume and on the quotient of the two 'time constants' for the start of an effect and its reversibility. Prevailing is the constant of reversibility (v. the linking of carbondioxide or oxygen to hemo-

globin). The size of this constant determines the type of a poison. If it
is small, the effect depends on concentration. If it is larger, a partial
accumulation of effects exists. If the constant approaches infinity,
i.e. if the effects are irreversible during the period of observation, the
effects are added. This happens, for instance, for cancer inducing agents.
The equations for irreversible summation are identical to well known
formulae of the 'theory of hits'. Completely separated phenomena
may be reduced to the same basic equation. This agreement supports
the hypothesis that all these processes are ruled by statistical laws.
They must be reducible to quanta and therefore could be described by
probability theory. But the equations and curves are trivial. No
conclusions are possible about the underlying elementary processes.

It is usual for pharmacological experiments that we do not observe
the primary effects, but only consequences which may be the results
of a long chain of consecutive reactions. Each step may be reversible
or irreversible. For the occurrence of a summation of effects it is
sufficient that a single step is irreversible. If two steps are irreversible,
an 'amplifying effect' exists which in principle corresponds to the
integral of concentration over time, multiplied by two.

Finally it is considered how the dosage-effect relationship depends
on the individual variation in mixed populations. It is emphasized
that according to experimental experiences the difference between the
sexes of a strain may be larger than that between two different strains.

H. GAUL, Voldagsen, and H. MUENZNER, Goettingen.
311 **Determination of the number of homologous chromosomes in
bastards of different species and subspecies.**

Problems on the homology of chromosomes in bastards of different
species or subspecies are theoretically important with respect to the
mutual relationship and the phylogeny of the parents which are used
in the crossing. They are essential also in practical breeding of plants.
Bastards of different species or subspecies show a variability of the
numbers of chiasms and bivalent chromosomes in the cells of the pollen.
Therefore it has not been possible yet to gain exact information on the
number of homologous chromosomes by making cytological observa-
tions. Empirically a parabolic dependence has been found between
the number B of fixed chromosomes and the number X of chiasms.
By using a combinatorial reasoning the same parabola is to be expected,
assuming that the chiasms are distributed randomly in the set of paired
chromosomal segments. The parameter of the parabola enables us to

estimate the number P of homologous chromosomes which are elegible for joining each other. Finally other models are tested with respect to their agreement with empirical findings.

312 F. KEITER, Hamburg. **Biometry and hereditary traits depending on many genes.**

A heredity which depends on many genes (better: on many factors since genes are not the only participants) is revealed by the variation of the trait in the population. Continuous, unimodal, symmetric variation is to be expected for the case of many collaborating genes, whereas discontinuous, asymmetrical variation corresponds to a single active gene. More than one mode may occur if a main gene and accessory genes participate, or if the influence of the environment is substantial. If the variation is plotted on the correct scale, traits depending on many genes prevail over those based on single genes, at least in normal anthropology.

In special heredity studies (parents-offspring comparison) the average of the children is found between the average of the parents and the mean of the total population. The variance is wide, only about 10% less than the variance of the population. The regression to the mean was only 15% for traits determined by impression, about 30% for measurements on adult offspring, about 45% for measurements on non-adult offspring. Mutually similar parents do not have more similar children than different parents. Evidently they are heterozygotes to the same degree. Children of certain combinations of parents have symmetrical, even normal distributions. The distribution stays symmetrical even for extreme values.

The same phenomena which belong to a polyfactorial heredity may occur with a single active gene if types of families exist in the population, representing a different heredity of the same trait. This is well known for hereditary diseases. There seems to be no possibility of separating the two cases.

Differences of the heredity of polyfactorial traits occur mainly because of a different regression to the average, less frequently because of a different variance of the children. The general scheme of the heredity of these traits, i.e. of almost all traits dealt with in normal anthropology, should be analogous to a high degree. This corresponds to the actual findings of critical values. For all possible combinations of child, mother, father the frequency for paternity is divided by the frequency without paternity. This ratio is the critical value. The critical values (proving values) are small for most single traits. Never-

theless they become very high for a combination of series of traits. For every polyfactorial trait there are combinations which exclude a paternity. Regions of variation exist which are impossible for children of certain combinations of parents. Usually the negative proofs are more convincing than the positive ones. A biometrically correct treatment of polyfactorial traits results in statements for diagnosing paternity which are as clear as those based on classic Mendelian heredity. Being an empirical hereditary prognosis, the method is free of hypotheses which are hard to verify.

313   P. KNEIP, Cologne. **Remarks on the evaluation of quantitative dosage-effect experiments.**

A series of tests is not completely evaluated if DL 50 and its variance have been determined. Further knowledge about drugs may be gained by plotting DL 50 against the duration of the experiment. This additional information does not depend on supplementary experimental animals. The method is simple and can be included in routine tests.

314   S. KOLLER, Wiesbaden. **Checking homogeneity if the regressions of several systems of correlation are analyzed.**

As an example of an analysis of covariances the regression lines in subsets of a large mass of data are compared with respect to their stability. These data belong to studies on the correlation between hemoglobin contents (in ccm blood) and surface area of erythrocytes (in ccm blood). In this example contradictions occur if one and the same relationship is assumed for men, women, and newborn infants. Checking the stability of a regression line in subsets of data corresponds under certain conditions to a test for the direction of the relationship. If actually X prevails over Y, the flat regression lines agree; if Y prevails over X, the steep regression lines are stable. It is assumed that no disturbing factors occur.

315   W. LUDWIG, Heidelberg. **Remarks on elementary problems which arise frequently in biometric routine work.**

As an introduction to the following "Discussion of Queries" elementary statistical problems are chosen which according to practical experiences arise again and again. An attempt is made to indicate

convenient methods which yield an accuracy in general sufficient for biological and medical research.

(1) Deletion of apparently extreme values of a small sample (normal distribution). (2) Guessing of a significant deviation from a normal distribution in small samples. (3) Separation of a non-normal distribution into two normal components if there is a hypothesis that the population is a mixture of two normal collectives. (4) The coefficient of variation. (5) Comparison of two means for a weak relationship (normal distribution). (6) The Brandt-Snedecor formula for very small samples. (7) Comparison of an empirical and a theoretical frequency or of two empirical frequencies for assumed binomial distribution and very small samples. (8) $2 \times 2 \times 2$-table and related topics.

316 W. LUDWIG, Heidelberg. **Stochastic reasoning in diagnosing paternity.**

A coefficient of plausibility Pl is defined that a defendant C_i , named by the mother C_i of the infant, is really the father of the suing child. The general concept of 'combination of degrees of traits (C)' is applied. Genetic and social-biological indications to paternity are separated. The result is a 'generalized and corrected Essen-Moeller formula'. At the same time statements are possible under which restricting assumptions the classic Essen-Moeller formula and other equations stay correct.

317 E. WALTER, Goettingen. **Components of covariance.**

The covariance may be split into components like the variance. The underlying model is described for the case of a simple classification. Sufficient methods for the computation of confidence intervals have not been developed yet, tests are lacking. Therefore the application of distribution-free procedures is discussed for a numerical example. This method may be used in animal husbandry for estimating the genetic correlation.

318 H. BAITSCH, Munich. **Biometry and problems of a correlation between traits.**

There are two main causes for a correlation of traits. One is the correlation following from a common causal (genetic) source. Then a complex of many traits is reduced to few arbitrary measurements. The other main cause for a correlation of traits is an inhomogeneity, an

incomplete mixture in the observed total population. Errors result from these correlations. In order to avoid them in a usual balancing procedure, a limitation to uncorrelated traits, if possible of a highly convincing kind, is recommended. Otherwise the various partial correlations have to be computed. The efficiency of the tested traits has to be reduced according to these partial correlations. Consequences of an incomplete mixture cannot be annulled by using such methods. Another solution can be found by applying a discriminant function instead of a balance. Consequences of a correlation of traits are—at least partially—eliminated automatically. Problems resulting from an incomplete mixture may be attacked more easily with these procedures.

Biometric Society (British Region) The twentieth meeting of the Region was held at the Wellcome Research Institute, 183 Euston Road, London, N.W. 1., at 2:30 p.m. on Wednesday, 14th April '55. The following papers were read and discussed:

319 P. ARMITAGE, A. W. DOWNIE and K. McCARTHY. Variations in counts of smallpox virus lesions.

When a suspension of smallpox virus is inoculated into a number of eggs, the variation in the count (i.e. number of lesions per egg) may be much higher than would be expected from a Poisson distribution. 92 groups of replicate counts were examined, and by working with log count and also using a logarithmic transformation of the variance of log count, a simple empirical relationship was established between the variance of the count and its mean ($\sigma^2 = 13.6 \ \mu$). There were no significant differences between groups in this relationship. It is proposed that, in comparing the means of two small groups of counts, the standard error of the difference could be estimated from this empirical formula, so as to provide a more powerful test than the t-test.

320 D. R. COX (Statistical Laboratory, University of Cambridge). The design of an experiment in which some treatment arrangements are inadmissible.

Consider an experiment in which the experimental units are arranged in *sets* of k units, a set corresponding, for example, to a single production run of an industrial process. Suppose also that the k units in each set are arranged in order corresponding to the first, second, etc. *period* of the set, and that for practical reasons there is a restriction on the order

of treatments within each set, such as that the level of the treatment must not decrease from one period to the next in a set. This paper is concerned with designs for such a situation; the method of construction is described and designs are given for a few special cases. Dr. C. J. Anson, G. K. N. Group Research Laboratory, suggested the problem; it arose in connection with an experiment on the properties of alloys made from high purity metals.

321 F. YATES: **The combination of data from a set of 2 × 2 tables.**

If a pair of treatments is such that their effects can only be measured by quantal ("all-or-nothing") responses the results of an experimental comparison of the two treatments can be arranged in the form of a 2 × 2 contingency table. When several such experiments are carried out direct pooling of the results can be misleading if there is heterogeneity between different experiments. In order to avoid pooling such data have often been analysed by calculating the significance level of each experiment separately and forming a combined significance test. This method, however, is inefficient, and also fails to provide a quantitative estimate of the difference between the treatments. A more satisfactory approach is to obtain a direct estimate of the difference (together with its standard error). If the numbers of observations in the separate experiments are small a maximum likelihood solution based on one of the well-known transformations (log log, logit or probit) should be used. The appropriate method of analysis will be described and illustrated by application to a genetical example. The method can easily be extended to sets of experiments involving more than two treatments.

The Biometric Society—British Region Wednesday, January 16, 1955

322 M. R. SAMPFORD. **The Use of Litter-Mates in Response-Time Experiments.**

(The use of litter-mates in comparative trials in which time to response is the observed variate is of considerable value in reducing error, but leads to complications in the analysis when some animals fail to respond before observation is suspended, or do not show the response at all. These two situations are discussed, and appropriate methods of analysis are outlined).

Abstracts of Papers presented before British Region on March 3, 1955

323 R. E. BLACKITH. **The Analysis of Social Facilitation at the Nest Entrances of Some Hymenoptera.**

The passage of unmarked social hymenoptera in and out of their nests is decisively non-random, grouping being demonstrated with both wasps and bumble bees. The observed distributions follow the negative binomial, one plausible interpretation of which assumes that workers are inhibited from passing through the nest entrance until sufficient individuals have accumulated to act as a releaser. Most workers of the red wasp *Vespula rufa* are released by the accumulation of from one to three further workers at the entrance. Other species of wasp seem to have less marked inhibitions. Young queen bumble-bees (*Bombus lapidarius*) have significantly higher inhibitions than have workers of this species. Worker wasps may obtain their releaser from individuals passing in the opposite direction only when insufficient pass in the same direction.

Different types of test reveal non-random passage of the nest entrance when many or when fewer insects are active. Grouping may be measured by the entropy of social organization. Some methods of estimating the number of workers foraging and of the mean duration of a flight, depend on a complete return of workers to the nest at night. A dawn to dusk record shows that this return may be far from complete, leading to biassed estimates.

324 CEDRIC A. B. SMITH. **An Estimation procedure for proportions, with genetical applications.**

Many parameters of genetical interest are the frequencies of particular types of events or objects: for example, gene frequencies, frequencies of recombination, "penetrance" or manifestation frequency, and so on. If we have a series of trials in each of which it is known whether the event in question has or has not occurred, or object been present, then the frequency is estimated as the proportion of such events in the whole sample, and the usual binomial formula gives the standard error. This applies, for example, to the estimation of the MN blood group gene frequencies by simple counting of genes in a sample of unrelated individuals. Complications are introduced by effects like dominance, which makes it uncertain exactly which genes are present (a group B individual can be genetically BB or OB), and by family data in which the same gene may recur among different members of the

same family. A counting method can still be used for estimation. Thus, in considering blood groups, we take provisional values of the gene frequencies, estimate from these how many B individuals are in fact BB, and how many OB, count genes, and so obtain improved estimates. An iteration leads to the final estimates, which (correctly calculated) can be shown to be Maximum likelihood estimates. However the process is purely numerical, avoiding the use of calculus. The variance of the estimates follows from suitably modified binomial or multinomial formulas, and the usual maximum likelihood theory can be applied to give heterogeneity tests, etc. The method is applicable whenever the probability of the observed sample is a rational function of the unknown parameters.

ERRATA—W. T. Federer and C. S. Schlottfeldt,

The Use of Covariance to Control Gradients in Experiments, June, 1954.

Gratitude is expressed to Prof. Gertrude M. Cox for pointing out some computational errors on Page 288 and 289, Volume 10, of the article entitled "The Use of Covariance to Control Gradients in Experiments." $b_{y1,2} = 0.198933$ should read $b_{y1,2} = 19.893256$. The corrected values for columns 5, 7 and 8 in Table VII are:

Adjustments for Total		Adjusted	
$b_{y1,2} (X_{1i.} - 0)$	$b_{y2,1} (X_{2i.} - 32)$	Total	Mean
−39.787		8755.975	1094.50
−59.680		8700.407	1087.55
99.466		8817.930	1102.24
59.680		7723.106	965.39
59.680		7527.906	940.99
−59.680		8168.024	1021.00
−59.680		7005.253	875.66
−.001		56698.601	. . . .
. . . .	. . . . . .	. . . .	1012.5

THE BIOMETRIC SOCIETY

General election. As general officers of the Society for 1955, The Council has re-elected Professor W. G. Cochran of Johns Hopkins University, President and C. I. Bliss, Secretary-Treasurer. In a total count of 454 individual mail ballots, the following were elected to the Council for 1955–57: G. M. Cox, B. B. Day, J. H. Gaddum, M. P. Geppert, M. Masuyama, P. A. Moran and J. Neyman. The Society is indebted for their services to the retiring Council members for 1952–54; C. W. Emmens, J. O. Irwin, Arthur Linder, A. M. Mood, C. R. Rao and Georges Teissier.

Biometric Symposium in Brazil. Plans are nearing completion for the International Biometric Symposium to be held in Campinas, near Sao Paulo, Brazil, on July 4–8, of this year. A preliminary announcement, dated April 19, has been sent to a special mailing list of nearly 300 in Latin America. A varied program, still provisional, has been arranged for the five days of the Symposium. The opening session will feature an address by W. G. Cochran, President of the Society. The Symposium will continue in the afternoon with two papers on Biometrical Genetics, by E. R. Dempster and by Sir Ronald Fisher. Experimental Designs for Perennial Crops and for Animal Experiments will be discussed on the following day by S. C. Pearce, C. Fraga and A. Conagin, G. M. Cox, F. Pimentel, P. G. Homeyer, W. J. Youden, and Arthur Linder. That evening Professor Th. Dobzhansky will lecture in Portuguese on "Genetica and Heterose". A session the following day on Medical Statistics will present papers by J. O. Irwin, J. Manceau, A. E. Brandt, and A. Vessereau. The rest of the day has been left free for excursions. On Thursday, different aspects of Sampling Techniques will be considered in the morning by M. H. Hanson, P. V. Sukhatme and V. G. Panse, E. Cansado, and J. Nieto de Pascual. A panel discussion on Experimental Designs is scheduled for that afternoon. The Friday sessions concern Bioassay, with papers in the morning by C. I. Bliss and by D. J. Finney, followed in the afternoon by a panel discussion on Statistical Problems in Bioassay submitted by those attending. Anyone interested in receiving announcements of the Symposium is invited to write to the Secretary of the Biometric Society, Box 1106, New Haven 4, Connecticut.

IUBS. At the 12th General Assembly of the International Union of Biological Sciences on April 12–16 in Rome, the Biometric Section of the IUBS, which is provided by the Society, was represented by L. L. Cavilli-Sforza of Milan and A. Vessereau of Paris. An additional

report has been received from A. Linder of Geneva, past President of the Society, who attended as Treasurer of the IUBS. During the Assembly, the IUBS was reorganized into three main divisions of Plant Biology, Animal Biology and General Biology, each with three to five sections. The Division of General Biology now comprises the Sections of Biometry, Cell Biology, Genetics, Microbiology and Limnology. Professor Linder resigned as Treasurer and was replaced by Dr. Lanjouw, a botanist from Utrecht, Holland. The President of the IUBS, Dr. Hörstadius of Sweden, commended the Biometric Section (Society) on different occasions as a model which could well be followed by others, in particular because of its international, regional and national organization. The Assembly approved support for an International Symposium to be held during our Fourth International Conference in Canada in 1958 on a biometric genetic topic, and will be able to give some financial assistance to the Secretary's office. Continuing support for a European Biometric Seminar or Colloquium, which it is proposed to hold annually in different parts of Europe, will depend largely upon the state of the budget of the IUBS. Although some controversy developed over IUBS support for this proposal, it was warmly endorsed by President Hörstadius and by other officers of the IUBS. Future subsidies have yet to be determined by the Executive Committee of the Union.

Biometric Colloquium in Italy. The European Seminar or Colloquium in Biometry, noted in BIOMETRICS for March, will be held at Varenna, Italy, on September 7–23, 1955. The following report is based upon the recent announcements issued by the Italian Region. The Seminar is open to graduates in medicine and surgery, in veterinary medicine, in the biological and other natural sciences, in agriculture, and in pharmacy, who wish to improve their knowledge of biometry for purposes either of teaching or of research. Three basic courses will be offered in Italian on (A) Fundamental Theory by M. P. Geppert of the W. G. Kerckhoff-Herzforschungs Institut, Bad Nauheim, Germany, (B) Design of Experiments by F. Anscombe of the Statistical Laboratory, University of Cambridge, England, and (C) Analysis of Variance and Covariance by C. A. B. Smith, Galton Laboratory, University College, London. Practical exercises in application will form part of the last two courses. Additional lectures have been arranged, both general and on specialized topics, including bioassay, animal husbandry, agricultural experiments, medicine and hygiene, and statistical genetics. Among the visiting professors on this part of the program are G. Barbensi, F. Brambilla, Sir Ronald Fisher, G. Pompilj and A. Tizzano. Problems submitted by participants in the Colloquium will be discussed in general seminars. An attendance of about 25 is anticipated. Applications for admission

to the Seminar and requests for further information should be sent at once to Professor L. L. Cavalli-Sforza, Istituto Sieraterapico Milanese S. Belfanti, Via Darwin 20, Milano, Italy, giving full information about the preparation of the applicant.

Syllabus on Biometry. During its Assembly in Rome, the IUBS sponsored a Symposium on "Problems of International Concern in the Life Sciences". A session on Education was chaired by Dr. Paul Weiss of the Rockefeller Institute for Medical Research in New York. At Dr. Weiss' request, a six-page mimeographed report on "Biometric Needs and Opportunities in Biological Education" was prepared in the Secretary's office. Based upon a statement by President Cochran, it was revised and expanded with the aid of 20 replies from members of the Society in the United States, Great Britain and Europe. The report reviews briefly the content and approach in a non-mathematical introductory course on the statistical aspect of biometry, the additional topics which might be considered in further or more specialized training, the place of laboratory work and conferences, the role of a statistician in biological research, and the place of refresher courses and of workshops or colloquia for the professional biologist. Members of the Society can obtain copies of this report on request from the Secretary's office.

German Region. The members of the Biometric Society in Germany held their third meeting and second Biometric Colloquy at the Kerckhoff-Institute in Bad Nauheim on January 28–30, 1955, with more than 120 persons in attendance, among them 40 members of the Society. The opening session on "Analysis of covariance" offered introductory reports by H. Münzner, S. Koller, C. Harte and E. Walter; the afternoon session on "Dose-response-curve" reports by R. Prigge, H. Druckrey, K. Soehring and K. Sommermeyer from the point of view of immunology, pharmacology and radiology. This topic was continued the second day with original papers by P. Kneip, A. Beckel and L. Schmetterer. The third day's program on "Biometric methods of paternity-diagnosis" consisted of papers by H. Gaul and H. Münzner, F. Keiter, H. Baitsch, W. Ludwig, W. Bauermeister and R. K. Bauer. On the second day a business meeting was followed by a discussion on "Unification of biometrical terminology (terms and symbols)". In a session on "Questions from practical biometric work", opened by a paper of W. Ludwig, 8 queries presented by the participants in the Colloquy were discussed at length.

During the Colloquium, the business meeting on January 29 discussed at length the organization of the German Region of the Society, voted to form the Region, adopted statutes, and fixed the Regional

dues. A later mail ballot named the following as Regional officers: President, E. Ullrich; Secretary-Treasurer, W. Ludwig; Regional Committee, K. Freudenberg, M. P. Geppert, O. Heinisch and H. Münzner.

On March 3–11, Professor R. C. Bose of the University of North Carolina gave a series of eight lectures on "Incomplete Block Designs" at the University of Frankfurt, to which all German members of the Society and other biometricians were invited on behalf of the Society and University. Dr. Bose's lectures were enthusiastically received and contributed materially to the development of biometry in Germany.

British Region. At the meeting of the British Region on April 14, 1954, at the Wellcome Research Institute in London the following papers were presented and discussed: "Variations in counts of smallpox virus lesions" by P. Armitage, A. W. Downie and K. McCarthy, "The design of an experiment in which some treatment arrangements are inadmissible" by D. R. Cox, and "The combination of data from a set of 2×2 tables" by F. Yates. On June 18, 1954, the Region met for dinner at the Lister Institute, which was followed by demonstrations of some of the work in progress.

The annual meeting of the British Region on January 26, 1955, elected the following Regional officers and committee: President, R. R. Race; Treasurer, A. R. G. Owen; Secretary, E. C. Fieller; Committee, D. J. Finney, M. J. R. Healy, J. A. Fraser Roberts, J. G. Skellam, J. M. Tanner, K. D. Tocher, J. W. Trevan, G. E. P. Box, and ex-officio Sir Ronald Fisher, J. H. Gaddum and F. Yates. Following the annual meeting, three papers were read and discussed: "An unusual frequency distribution" by Sir Ronald Fisher, "Estimation of bacteria in whale meat by dilution methods" by H. W. Daniels, and "The use of litter-mates in response-time experiments" by M. R. Sampford. The Region met again on March 3 at the Wellcome Research Institute in London, with the following program: "Analysis of social facilitation at the nest entrances of some Hymenoptera" by R. E. Blackith, "An estimation procedure for proportions, with genetical applications" by C. A. B. Smith, and "Trials of skinfold calipers" by M. J. R. Healy and J. M. Tanner. Abstracts are being published in BIOMETRICS.

ENAR. The Eastern North American Region met jointly with the Institute of Mathematical Statistics on April 22–23 at the University of North Carolina in Chapel Hill. At the opening session, invited papers by F. S. McFeely, J. E. Freund, T. Horner, I. Miller, H. Bozivich, and R. L. Wine considered various aspects of life testing, components of variance and decision procedures. At the following session D. G. Austin, J. Blackman and C. Derman spoke on Probability Theory.

The afternoon program opened with a session on Multivariate Analysis, with papers by T. W. Anderson, W. G. Howe and H. C. Sweeny, which was followed by nine contributed papers on mathematical statistics. A discussion of the relation between smoking and mortality from lung cancer, opened the meeting on April 23 with J. Cornfield and W. Haenszel as principal speakers, and B. Harshbarger and D. Horn as discussants. The morning program concluded with a session of five papers on mathematical statistics. The afternoon program included contributed papers on problems in discriminant analysis by S. W. Greenhouse, in bioassay by J. Ipsen, in plot selection by D. G. Horvitz and J. Fleischer, in characteristic functions by M. C. K. Tweedie, and by title, on contingency tables with missing or mixed-up cell entries by G. S. Watson.

Abstracts of the Society sessions are printed in this issue of Biometrics; those of the joint sessions and others will appear in the *Annals of Mathematical Statistics*.

Région Française. Lors de la dernière réunion de la Société Française de Biométrie, qui eut lieu mercredi le 9 février à l'Ecole Normale Supérieure à Paris, Monsieur S. Lédermann fit une conférence sur "le Cancer, l'Alcool, et le Tabac" et Messieurs J. Sutter et L. Tabah discutèrent les "Recherches sur la Mortalité par vieillissement". Au cours de cette réunion eut lieu l'élection pour le renouvellement du Couseil et du Bureau. Monsieur David Schwartz fut élu secrétaire-trésorier et Monsieur J. M. Faverge fut élu membre du conseil.

EXPERIMENTAL DESIGN IN INDUSTRY*

H. C. Hamaker

Philips Research Laboratories
N.V. Philips' Gloeilampenfabrieken
Eindhoven, Netherlands

1. *Introduction*

The *design of experiment* and *analysis of variance* are techniques which have been developed mainly in connection with agricultural research. The majority of examples in textbooks on the subject are consequently drawn from this field.

These techniques also apply to industrial and technological experimentation. The main purpose of this note is to emphasize that the conditions under which these techniques have to be applied in industry are in several respects essentially different from those prevailing in agriculture; and that these differences imply changes in the method of teaching, of analysis, and of presentation if we are to reap the full benefit and efficiency of these statistical methods in the industrial field.

2. *The difference in speed and its consequences*

One of the main differences is a difference in speed. The agriculturalist is usually restricted to one experiment in a year. Hence he has ample opportunity to plan his experiment, to carry it out, and to analyze it before the planning of the next experiment is started. If we are limited to one experiment per season it is evidently of great value to set up a fairly complex experiment so that the maximum of information can be collected in a reasonable time. It will likewise pay to have the entire procedure supervised by a staff of competent scientists with a university background.

Not so in industry. A single machine produces 1200 light bulbs or radio valves in one hour; and even a slower product such as the wheel of a railway carriage is shaped within about 10 minutes. Also, where mass production is going on on a vast scale and is consequently

---

organized as a routine, its supervision is not as a rule in the hands of scientists but is entrusted to personnel with a secondary-school education, maybe with some additional technical training.

Carrying out a designed experiment under those conditions requires a considerable effort in organization, and production managers will only be inclined to undertake such methods of experimentation if they are convinced of their utility and if they are able to comprehend the useful results achieved.

In the first place this requires a not too complex set-up of the experiments; two- or three-way classifications and latin squares are already an important improvement as compared to the one-way classification of classical laboratory experiments. Owing to the high speed such simple experiments can easily be repeated with slight alterations; and more complex experiments, which are difficult to explain and may aim at too much information at once, can often be avoided.

Another and very important requirement is that we must explain our purposes in a language that factory people will understand, that is, by numerical example. We have all learned from early childhood to reason in figures, rather than in symbols and abstract formulae. Production managers in particular, whether they are statistically minded or not, look daily at figures representing their stock of raw materials, the amount and quality of items produced, the scrap discarded, etc. They will consequently grasp the meaning of a designed experiment much sooner and easier from a numerical example than from a symbolic model, provided the numerical data are presented in a form which corresponds as closely as possible to the producer's technological experience.

We will illustrate this by some numerical examples which all refer to the simplest type of designed experiment, viz. the two-way classification.

3. *Some industrial examples of two-way classifications*

Five nickel rods of 1 mm diameter are put in a metallic clamp, jointly immersed in a suspension of aluminum oxide (fig. 1), and for a few seconds a tension of some 100 volts is applied between the nickel rods and the electrically conducting vessel containing the suspension, the rods being negative. Since the oxide particles carry a positive charge owing to their colloid properties, they move towards the nickel electrodes and are deposited there as an oxide coating. To investigate how the amount deposited varies with position and height, the thickness of the coating was observed in three heights ($H_1 - H_3$) on each of the 5 rods (positions $P_1 - P_5$), the results being as recorded in table 1.

TABLE 1

Thickness of an aluminum oxide layer observed at 3 heights on 5 nickel rods

				Position of Ni rod			
		P_1	P_2	P_3	P_4	P_5	
				Z_{ij} in microns			$Z_i.$
	H_1	125	130	128	134	143	132
Height	H_2	126	150	127	124	118	129
	H_3	130	155	168	159	138	150
	$Z._j$	127	145	141	139	133	137

TABLE 2

Analysis of variance of the data of table 1

Source	S.S.	D.F.	M.S.
Positions P	600	4	150 μ^2
Heights H	1290	2	645
Residual	1168	8	146

If we analyze these data in the usual way (table 2) we find significant differences between heights but no significant differences between positions. The trouble is, however, that if we present the analysis in this form to people in the factory, where the technique is used for coating the heating filaments of radio-valve cathodes, it will not be understood. Reading and interpreting sums of squares, degrees of freedom, and mean squares requires a sophisticated statistical training which we do not encounter in an industrial environment. Besides, if we accumulate a sufficient number of observations even the smallest differences will eventually become statistically significant, but this statistical significance tells us nothing whatsoever as to their technological significance. An effect that is statistically significant may be technologically quite negligible; and conversely statistical insignificance does not disprove the presence of effects that may be technologically worth consideration.

Hence for industrial purposes we must seek a method of presentation which is more easily understood. We have found the method explained in table 3 very useful.
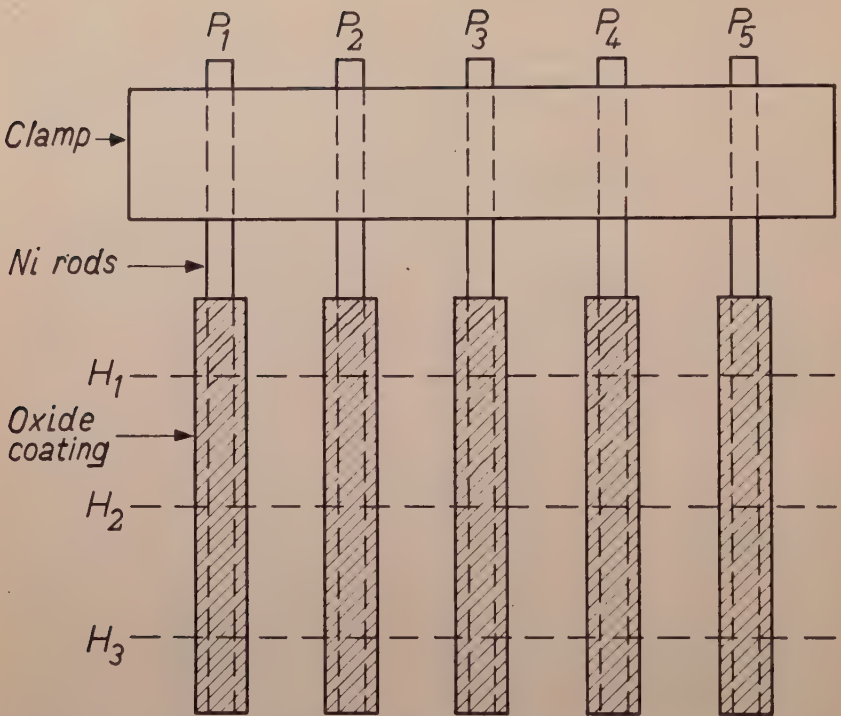
FIG. 1.   FIVE NICKEL RODS SIMULTANEOUSLY COATED WITH ALUMINUM OXIDE
LAYER BY ELECTROPHORESIS.

TABLE 3

A simple and useful presentation of the analysis of the data in table 1

Average	Positions		Heights		Residual
137 μ	P_1 $-$ 10 μ P_2 $+$ 8 P_3 $+$ 4 P_4 $+$ 2 P_5 $-$ 4		H_1 $-$ 5 μ H_2 $-$ 8 H_3 $+$ 13		$s = 12 \mu$; $\nu = 8$
s s'	$12/\sqrt{3}$	7.1 μ; $\nu = 4$ $= 6.9 \mu$; $\nu = 8$	$12/\sqrt{5}$	11.3 μ; $\nu = 2$ $= 5.4 \mu$; $\nu = 8$	

The general average is 137μ and we may express the influence of
height and position by computing the difference between row and
column averages from the general average.   These data are given in
table 3; from them we see that the average layer thickness in position

1 is 10μ below the general average and for height 3 it is 13μ above. These are figures from which the man in the factory can judge the technological importance of the effects observed.

From these data we may now proceed to predict that the average thickness in position 2 and height 3 will be

$$137 + 8 + 13 = 158\mu, \tag{1}$$

and by carrying out the same computation for all positions and heights we find that we can explain the part of the observations recorded in table 4; and by subtracting from the original data we find what portion is still left unexplained.

TABLE 4

The parts of the data of table 1 that are explained and unexplained by the simple additive formula (1).

Part explained					Unexplained				
122	140	136	134	128	$+3$	-10	-8	0	$+15$
119	137	133	131	125	$+7$	$+13$	-6	-7	-7
140	158	154	152	146	-10	-3	$+14$	$+7$	-8

Clearly equation (1) assumes that positions and heights act independently so that their effects may simply be added. The unexplained part of the observations may partly be due to this assumption being too simplistic, partly they may result from random fluctuations which can never be avoided. If we interpret the unexplained part as entirely due to random errors we can proceed to estimate the standard deviation by dividing the sum of squares by the number of degrees of freedom; since the sums of the elements of the unexplained part of the observations are zero both in a horizontal and in a vertical direction it is easily seen that all elements are determined when the eight elements within the dotted frame are fixed; hence the number of degrees of freedom is 8.

From the estimate of error, $s = 12\mu$, $\nu = 8$, thus obtained we may predict a standard deviation between positions of $12/\sqrt{3} = 6.9\mu$, since the average for each position is based on 3 observations. The actual standard deviation computed from the position averages is 7.1μ, $\nu = 4$ and this is clearly not significant. The same argument applied to heights yields a predicted standard deviation of $12/\sqrt{5} = 5.4\mu$

TABLE 5

Fluidity of iron as a function of Si-content for three replicates

	1.25	1.50	1.75	2.00	2.25	% Si
1	47.5	60.0	65.0	72.5	77.5	
Replication 2	55.0	55.0	67.5	75.0	85.0	
3	37.5	50.0	70.0	75.0	75.0	

TABLE 6

Various methods of applying analysis of variance to the data of table 5

	Source	S.S.	D.F.	M.S.
I	Treatments	2177	4	544
	Residual	275	10	28

	Source	S.S.	D.F.	M.S.
II	Linear term	2125	1	2125
	Residual	326	13	25

	Source		S.S.	D.F.	M.S.
III	Linear term	L	2125	1	2125
	Quadratic term	Q	33	1	33
	Replications	R	90	2	45
	$R \times L$		40	2	20
	$R \times Q$		108	2	54
	Residual		55	6	9

against a computed value of 11.3μ, $\nu = 2$, and the appropriate statistical test shows that this difference is significant at about 5% level.

Of course the entire argument is essentially that of the analysis of variance, but it presents the results in a form that will be much more easily understood. From the differences between positions and between heights as given in table 3 the industrial technician will at once realize the technological importance of the effects observed. Also standard deviations, expressed in a dimension with which factory operators are

familiar and directly associated with the 2σ and 3σ limit concepts, are more easily understood. Variances must be considered as part of a statistical jargon which should preferably be avoided.

As stated before we have found this method of presentation most useful, because by it people will easily grasp the meaning of the statistical analysis and will be induced to apply it to good purpose to their own observations. Pretty soon, however, they will come across cases where a simple analysis by rows and columns does not give the full answer. The observations recorded in table 5 from a paper by A. Palazzi[1] are a case in point.
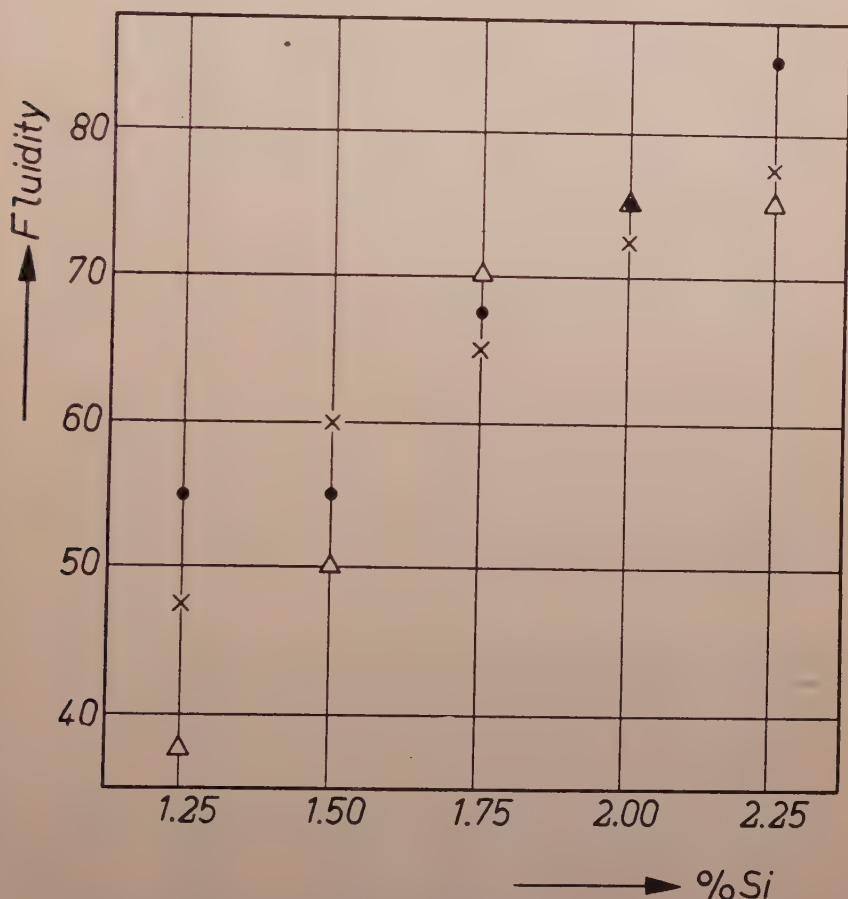
FIG. 2. FLUIDITY OF IRON PLOTTED AGAINST SILICON CONTENT; THREE REPLICATES.

The fluidity of iron was observed at 5 different levels of Si-content, in three replications (table 5). If we plot these observations (fig. 2) we observe a linear dependence of the fluidity on Si-content and an analysis in treatments and error is clearly unsatisfactory because this fact is not taken into account. Indeed if we introduce a linear component we find that nearly the full sum of squares which is in the first analysis attributed to 4 degrees of freedom for treatments is concentrated into one degree of freedom for regression in the second case; thereby the degree of significance is enormously enhanced (see table 6, I and II).

If we are so inclined the residual component can be split up into a number of parts testing separately differences between replications, between regression coefficients for the three replications, etc.; but this does not reveal any pronounced effects, as might have been surmised from the plot (table 6, III).

Finally, for presentation to the factory, sums of squares and mean squares are again unsuitable; the actual equation giving the linear relationship between the fluidity and Si-content plus one estimate of error is technologically much more convenient. We then arrive at the result:

$$
\left.
\begin{aligned}
F = 64.5 + 33.8(\text{Si}\% - 1.75) + e; \\
s(e) = 5.0; \qquad \nu = 13; \\
s(64.5) = 1.3; \\
s(33.8) = 3.7
\end{aligned}
\right\}
\tag{2}
$$

where $s(64.5)$ and $s(33.8)$ represent the standard errors in the constants 64.5 and 33.8 of the regression equation as estimated from the residual error $s(e) = 5.0$.

TABLE 7

Self-inductance of coils with iron-oxide cores at varying temperatures of the bridge; the recorded data are % deviations from a standard.

Temp. °C	Coil				
	1	2	3	4	5
21	1.400	0.246	0.478	1.010	0.629 %
23	1.400	0.235	0.467	0.990	0.620
24	1.375	0.212	0.444	0.968	0.495
25	1.370	0.208	0.440	0.967	0.495

A third somewhat more complex case is shown in table 7. Here the selfinductance of some coils with an iron-oxide core were measured while the temperature of the measuring bridge was varied, the coil temperature being kept constant; % deviations from a standard were observed.

When we plot the data (fig. 3) we see that with coil 5 some reading, or clerical, errors have occurred. This is a characteristic of industrial conditions. Clerical errors may always be expected and it is a useful, if not a necessary, practice to plot the data before a numerical analysis is started. In a treatment by purely numerical techniques clerical errors may easily pass unnoticed and cause serious distortions in the conclusions drawn.

Discarding the observations on coil 5 as evidently unreliable, we see for the remaining 4 coils a linear temperature dependence and deviations from a straight regression line which seem to possess the same sign and magnitude for one temperature independently of the coil measured. It might of course be possible to describe the effects by means of a third-degree equation in the temperature but physically such a complex relationship in so narrow a temperature range is highly unlikely.

In the experiment the same set of 5 coils were measured first early in the morning, and then at different times during the day while the temperature in the room was steadily rising. Each time a set of measurements was made, the zero-point of the apparatus had to be adjusted
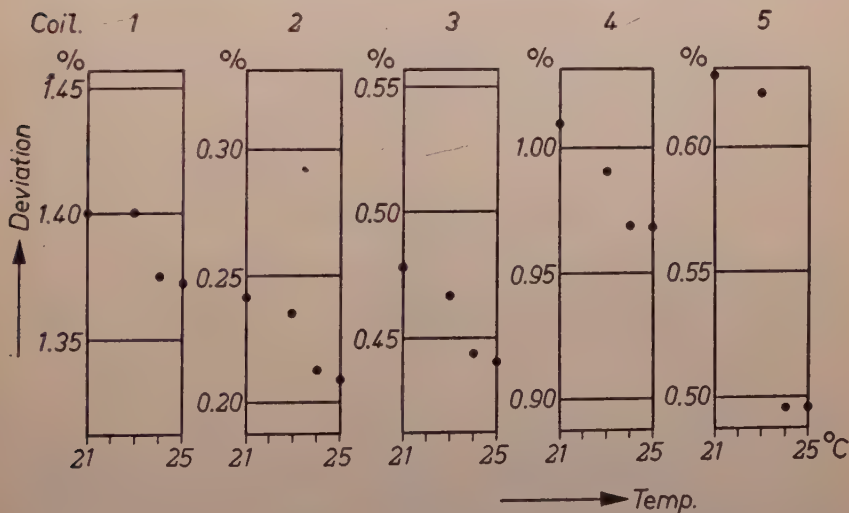
FIG. 3. SELFINDUCTANCES OF 5 COILS. DEVIATION FROM A STANDARD PLOTTED AGAINST THE TEMPERATURE OF THE MEASURING BRIDGE.

afresh. Also the temperatures recorded were rounded off to whole degrees and there may be errors in the temperature of up to 0.5°C. Both the zero-point adjustments and the errors in temperature will introduce errors in the selfinductances which vary in a random manner between the 4 sets of observations, that is, between temperatures in table 7 and in fig. 3, but are constant over the 5 coils at a given temperature. We shall call these errors the *errors of adjustment*.

On the basis of this argument it seems reasonable to assume the model

$$Z_{ij} = a + b_i + c_i T_j + \delta_j + \epsilon_{ij} , \qquad (3)$$

b_i specifying differences in the general level between coils,
c_i specifying temperature coefficients which may vary from coil to coil,
δ_j specifying the errors of adjustment, and
ϵ_{ij} the errors of observation.

The greek symbols δ and ϵ have been used to distinguish the random variables from the systematic phenomena denoted by roman letters.

The analysis of variance now runs as follows. First of all we may sum the observations on coils 1 to 4 for each temperature and carry out a linear regression analysis on the totals; this gives the sums of squares in the second row of table 8. Next we adjust a linear equation to each coil separately which gives the sums of squares in the first row of table 8. The last row contains the difference of these sums of squares.

TABLE 8

Steps in the analysis of variance of the data in Table 7 (coil 5 excepted) according to the model (3)

	Sum of squares for			
	Linear Regression	D.F.	Residual	D.F.
For coils individually	0.003583	4	0.000440	8
For sum over coils	0.003530	1	0.000376	2
Difference	0.000053	3	0.000064	6

In table 8 the total sum of squares, 0.003583, for linear regression has been split up into the amount 0.003530 with one degree of freedom for the regression common to all four coils and the amount 0.000053 with 3 degrees of freedom corresponding to differences in the regression coefficients between coils.

This well-known technique now also applies to the residuals. The sum of squares 0.000376 with $\nu = 2$ corresponds to residuals common to all coils at a given temperature; that is, to the random component δ in model (3). And the remaining sum of 0.000064 with $\nu = 6$ corresponds to the variation in the residuals from coil to coil; that is, with the errors of observations ϵ in model (3).

Apart from these effects we also have evident differences in the general level from coil to coil, so that the final analysis of variance takes the form of table 9.

TABLE 9

Analysis of variance of the data of table 7 in keeping with the model (3).

Source	S.S.	D.F.	M.S.
Linear component L	0.003530	1	0.003530
Residual I (δ)	0.000376	2	0.000188
Between coils C	3.28	3	1.09
Coils × Linear Comp. $C \times L$	0.000053	3	0.000018
Residual II (ϵ)	0.000064	6	0.000011

The errors of adjustment δ are the same for all 4 coils at one temperature. Hence these errors will influence the regression L common to all coils, but they will not influence the differences between coils or the $C \times L$ interaction. These last two effects must therefore be tested against Residual II, but the linear effects L must be tested against Residual I.

The analysis shows that we have no reason to assume differences in the regression coefficients between coils. The model finally leads to the following numerical results.

Coil   Changes Z in selfinductance in %

$$
\begin{array}{ll}
1 & Z = 1.389 \\
2 & = 0.228 \\
3 & = 0.460 \\
4 & = 0.986
\end{array}
\left. \right\} - 0.0100(T - 23)\% \quad (4)
$$

Reading errors $s_\delta = 0.0033\%$; $\nu = 6$

Errors of adjustment $s_\delta = 0.007\%$; $(\nu \sim 2)$

The estimate of the reading errors is simply the root of Residual II in table 9. To find an estimate for the errors of adjustment we must

recall that the Residual I in table 9 was obtained from the sum of 4 observations, one for each coil.   Hence the expected value of Residual I is

$$4\sigma_\delta^2 + \sigma_\epsilon^2 ,$$

and an estimate of σ_δ^2 is consequently given by

$$s_\delta^2 = \frac{0.000188 - 0.000011}{4} = 0.000044.$$

The correction for s_ϵ^2 is very small so that we may approximately assign 2 degrees of freedom to the estimate s_δ .

4. *The theory of two-way classifications*

The model usually applied to a two-way classification is

$$Z_{ij} = a + b_i + c_j + \epsilon_{ij} . \tag{5}$$

Above we have, however, encountered several examples where this model is inadequate and a further extension of the theory seems appropriate.

First of all we must distinguish between classifications according to *known* and *unknown* levels*.

If we wish to investigate differences in electron emission of radio valves when the cathodes have been made from different batches of raw materials we have a classification by unknown levels.   If, however, these raw materials have been chemically analyzed and classified according to their content of reducing impurities we have a classification according to known levels.   In the first instance we only ask whether there are differences between batches, while in the second instance we try to relate these differences to a measured chemical characteristic of the batches.

Likewise in the example of table 7 the classification by temperature is by known levels and the classification by coils is by unknown levels.

We may now further distinguish between three different kinds of two-way classifications:

    (A) unknown levels in both directions;
    (B) known levels in one direction;
    (C) known levels in both directions.

Different models must be applied to each of these situations.

---

*O. L. Davies and his colleagues[2] have introduced a distinction between *quantitative* and *qualitative* factors.   This paper was prepared before their book was available and it may be interesting to note that we have both independently felt the need for this distinction.

5. *Two-way classifications with unknown levels in both directions*

This is the case most commonly considered in textbooks and analyzed into a mean square for rows, for columns, and for residual. The model customarily assigned to this design is given by (5). Here we would, however, like to propose a change, replacing (5) by

$$Z_{ij} = a + b_i + c_j + d_{ij} + \epsilon_{ij} \tag{6}$$

where d_{ij} represents the interaction, that is, a systematic component depending both on row and column and due to non-additivity of the row and column effects, while ϵ_{ij} is the purely random component; a model that was also used by Tukey[3].

In industry one sometimes has to analyze two-way experiments which have been performed without replication. It must then be borne in mind that the residual contains the pooled effect of interaction and random fluctuations. It is an experiment in which these two components are confounded. The statistical significance will in such cases be underestimated because the estimate of error is biased in the direction of higher values by the interaction term.

Whether or not it is desirable to separate interaction from the random fluctuations depends on the situation envisaged. If we find pronounced and highly significant differences between rows and columns it will often be of small interest to know whether there is interaction, because this interaction is evidently small compared with other effects present and hence is of no technical importance.

A separation of the interaction and the random component can of course be carried out by replication, because in replications the interaction is repeated while the random elements change from one replicate to another.

A clear distinction between interaction and random components is, we believe, essential for a correct understanding of more complex cases.

In this connection a definition of the term *replication* would also seem desirable. In textbooks this term is often introduced without further explanation. To judge from its use in statistical analyses, *replication* is equivalent to the introduction of an additional factor of which we are convinced that it does not interact with the other factors of the experiment.

Replication on different blocks in one field satisfies this definition. We expect differences between blocks but no block-treatment interactions; these never occur in the analysis.

But in industry replications often mean repetitions of the experiment on different days. In the meantime conditions may have changed, say,

by changes in raw materials, by exhaustion of a chemical solution or otherwise, and we can never be quite sure that such changes do not influence the outcome of an experiment otherwise than by the general level only. We have occasionally observed interactions where they were not suspected at all. It may sometimes be wise to verify by an appropriate analysis whether the replications can be truly considered as such; and the term should not be used too loosely and should be properly defined.

An alternative technique for discovering interaction without replication will be discussed in §11.

6. *Two-way classification with known levels in one direction*

The examples presented in tables 5 and 7 are cases in point. From the accompanying discussion it will be clear that we now have the choice between a variety of models. If Y_j $(j = 1, \cdots n)$ signify the known levels for the n columns we can introduce a linear component cY_j into our model common to all the rows, or alternatively linear components c_iY_j with different higher-order components can be introduced in similar fashion.

As a rule, however, the variations in the levels are comparatively small and we need not go to higher powers than the second. As a matter of fact cases with a linear component occur quite frequently; cases where a significant quadratic component is observed are occasionally encountered in the literature, but we are not aware of any case where a cubic term was required; these are anyhow very rare.

As a general model for the design under consideration we may therefore pose

$$Z_{ij} = a + b_i + c_iY_j + d_iY_j^2 + \epsilon_{ij} . \qquad (7)$$

In some cases, as in table 7, it may be necessary to add a term η_j to account for random errors which are constant through each column; in that case the columns are classified by a mixture of known and unknown levels. In how far the model (7) can be simplified by omitting the index i from b, c or d, or by dropping the quadratic term will depend on the situation.

The customary procedure to treat a two-way classification of the present type is by first analyzing into rows and columns and by introducing a linear or quadratic component only when significant differences between columns have been found.

That this procedure is not quite satisfactory is demonstrated by the example of table 5. There an analysis into columns gave a mean square of 544 for 4 degrees of freedom, while the introduction of a linear component gave a mean square of 2125 for one degree of freedom and a

greatly enhanced level of significance. It may therefore well happen that we do not find significant differences between columns, where we would find a clearly significant linear component.

Hence the linear component should be introduced into the model right at the outset.

7. *The two-way classification with known levels in both directions*

In this situation the method of the previous section can be further extended. If X_i $(i = 1, 2, \cdots, m)$ and Y_j $(j = 1, 2, \cdots, n)$ denote the known levels for the m rows and n columns respectively, we can introduce not only terms in X, X^2, Y, and Y^2 into our model but also mixed or interaction terms of the form XY, X^2Y, etc. The full model up to the second degree becomes

$$Z_{ij} = a_{00} + a_{10}X_i + a_{01}Y_j + a_{20}X_i^2 + a_{11}X_iY_j + a_{02}Y_j^2 + \epsilon_{ij} . \tag{8}$$

The most convenient way to apply such a model is by means of orthogonal polynomials*. Let $\xi_k(X_i)$ $(k = 0, \cdots, m - 1)$ be the m orthogonal polynomials of degree k in X_i, and $\eta_l(Y_j)$ $(l = 0, \cdots, n - 1)$ similarly the n polynomials in Y_j of degree l; then the products

$$\xi_k(X_i)\eta_l(Y_j) \tag{9}$$

define mn two dimensional polynomials in X and Y which are mutually orthogonal over the complete set of mn combinations of X_i, Y_j. The regression coefficients are

$$b_{kl} = \frac{\sum_{ij} \xi_k(X_i)\eta_l(Y_j)Z_{ij}}{\sum_{ij} \{\xi_k(X_i)\eta_l(Y_j)\}^2} , \tag{10}$$

and the corresponding sums of squares

$$V_{kl} = \frac{\{\sum_{ij} \xi_k(X_i)\eta_l(Y_j)Z_{ij}\}^2}{\sum_{ij} \{\xi_k(X_i)\eta_l(Y_j)\}^2} . \tag{11}$$

In principle we can thus split the total sum of squares into mn different portions** each with one degree of freedom; as stated it will in practice

---

*The use of orthogonal polynomials for an analysis in two or more dimensions, though not common, is not new. The technique is, for example, described by DeLury.[4]

**Usually the total sum of squares $\sum_i \sum_j (Z_{ij} - Z_{..})^2$ is associated with $(mn - 1)$ degrees of freedom, the one degree of freedom for the general average $Z_{..}$ being disregarded because it is not of interest. In the present instance it is theoretically more convenient to retain this degree of freedom as part of our general argument. It might be used to test whether the average $Z_{..}$ differs significantly from zero, but in most situations such a test is quite superfluous. For the same reason it is appropriate to consider polynomials of degree zero as part of the complete set of orthogonal polynomials; they are of course constants and may be taken equal to unity: $\xi_0 = \eta_0 = 1$.

seldom be necessary to go beyond terms of the second degree. As an example of this type of analysis let us consider the data in table 10, giving the relative changes observed in a life test with resistors of different resistances and nominal wattages.

TABLE 10

Relative changes in resistance in per mil; each value is an average observed on 5 resistors.

Wattage	Resistance					
	100	200	500	1000	2000 Ω	
	Changes in resistance $= Z_{ij}$					$Z_{\cdot j} =$
1/8	2.8	2.6	3.0	2.5	2.1	2.6
1/4	1.9	2.1	4.1	2.2	1.7	2.4
1/2	3.2	2.8	1.9	3.4	4.2	3.1
1	2.0	3.4	3.1	3.6	3.9	3.2
2	2.1	2.6	3.4	4.3	6.1	3.7
$Z_{i\cdot} =$	2.4	2.7	3.1	3.2	3.6	3.0

An ordinary analysis of variance yields the following result.

TABLE 11

Ordinary analysis of variance of the data in table 10

Source	S.S.	D.F.	M.S.
Resistances	4.30	4	1.07
Wattages	5.30	4	1.32
Residual	14.88	16	0.93

There are no clearly significant effects. If we compute the residuals term by term, however, we obtain the results given in table 12.

TABLE 12

Residuals of table 10, when row and column differences have been eliminated.

	100	200	500	1000	2000 Ω
1/8W	+0.8	+0.3	+0.3	+0.3	−1.1
1/4	+0.1	0.0	+1.6	−0.4	−1.3
1/2	+0.7	0.0	−1.3	+0.1	+0.5
1	−0.6	+0.5	−0.2	+0.2	+0.1
2	−1.0	−0.8	−0.4	+0.4	+1.8

These residual terms indicate that the analysis of table 11 is not quite adequate; for positive and negative residuals are clearly concentrated along the two diagonals of the table, which cannot easily be understood if they are the result from purely random variations.

The values of the wattages are in geometrical progression and the values of the resistances very nearly so; it seems therefore reasonable to specify the levels of rows and columns by log W and log R respectively. We then have equispaced levels which can be represented by $X_i = -2$, $-1, 0, +1, +2$, and $Y_i = -2, -1, 0, +1, +2$, and we may use the existing tables of orthogonal polynomials[4,5]. Carrying out the analysis on this basis up to the second degree we obtain the result of table 13.

TABLE 13

Analysis of variance of the data of table 10 by means of orthogonal polynomials

Source	Polynomial	Leading term	S.S.	D.F.	M.S.
General average	$\xi_0\eta_0$	1	225.	1	225.
Wattages	$\xi_1\eta_0$	X	4.50	1	4.50
	$\xi_2\eta_0$	X^2	0.23	1	0.23
Resistances	$\xi_0\eta_1$	Y	4.50	1	4.50
	$\xi_0\eta_2$	Y^2	0.003	1	0.003
Interaction	$\xi_1\eta_1$	XY	7.13	1	7.13
Residual			8.14	19	0.44

Whereas the differences between rows and columns were not significant in the analysis of table 11, we now find linear terms in both directions which are very clearly significant. In addition we find an interaction term which is also highly important; by removing this component the residual mean square has been halved. We conclude that the changes in resistance values can be represented by a simple equation comprising terms in X, Y, and XY, the terms in X^2 and Y^2 being clearly unimportant. Numerical computation yields the equation:

$$R = 3.0\xi_0\eta_0 + 0.30\xi_1\eta_0 + 0.29\xi_0\eta_1 + 0.267\xi_1\eta_1 \text{ per mil} \quad (12)$$

or

$$R = 3.0 + 0.30X + 0.29Y + 0.267XY \text{ per mil,}$$

and the residual standard deviation is $s = 0.63$ per mil. If we calculate the expected changes from this equation and subtract these from the original observations we obtain the residuals entered in table 14.

TABLE 14

Residuals of the data of table 10 with respect to the model 12

	100	200	500	1000	2000 Ω
1/8 W	−0.09	−0.04	+0.60	+0.34	+0.19
1/4	−0.75	−0.58	+1.40	−0.52	−1.05
1/2	+0.78	+0.09	−1.10	+0.11	+0.62
1	−0.19	+0.66	−0.20	−0.26	−0.51
2	+0.15	−0.18	+0.20	−0.12	+0.85

The systematic distribution of signs noted in table 12 has now disappeared.

Of course the assumption that the levels for rows and columns are proportional to log W and log R is rather an arbitrary one. It leads to a satisfactory description of the data and from the point of view of the user of the resistors this may be considered a sufficient a posteriori justification. When, however, we are investigating the physical or chemical processes causing the changes in resistance the problem is quite a different one. It must then be borne in mind that there may be other ways of specifying the levels which lead to equally satisfactory presentations of the observations but are to be preferred in view of our concepts of the underlying processes.

Since analyses of this type are not very common a second example may be appropriate. It is provided by the observations on the thickness of aluminum-oxide layers recorded in table 1 and is of interest since it is not in this case the XY interaction which is of importance.

To treat heights H and positions P as unknown levels, as we did in tables 2 and 3, is not really satisfactory and it seems preferable to introduce known levels to describe the dependence of layer thickness on these two factors. Since both heights and positions are equally spaced we may again use the tables of orthogonal polynomials. Table 15 gives the final result of such an analysis.

<div align="center">TABLE 15</div>

<div align="center">Analysis of variance of the data of table 1 by orthogonal polynomials</div>

Source	Polynomial	Leading term	S.S.	D.F.	M.S.
Heights	$\xi_1\eta_0$	X	$810\,\mu^2$	1	$810\,\mu^2$
	$\xi_2\eta_0$	X^2	480	1	480
Positions	$\xi_0\eta_2$	Y^2	453	1	453
Interactions	$\xi_1\eta_2$	XY^2	603	1	603
	$\xi_2\eta_1$	X^2Y	346	1	$346\big\}$
Residual			366	9	$41\big\}\ 71;\ \nu = 10$

By the symmetry of the positions the absence of $Y(\xi_0\eta_1)$ can be immediately understood. And the significant XY^2 term $(\xi_1\eta_2)$ is explained if we assume that the linear effect with height for the two outer positions (P_1 and P_5) differs from that in the centre. Though the $X^2Y(\xi_2\eta_1)$ term nearly reaches the 1% level of significance, an effect of this type is technically difficult to explain. For this reason one may be inclined to disregard this term and pool it with the residual. If we do so the adjusted model is

$$Z = 137 + 9.0\xi_1\eta_0 + 4.0\xi_2\eta_0 - 3.3\xi_0\eta_2 - 4.6\xi_1\eta_2 \qquad (13)$$

and the residuals with respect to this model are those of table 16. That this model is not quite adequate is revealed by a somewhat systematic distribution of signs. This could be remedied by including a $\xi_2\eta_1$

<div align="center">TABLE 16</div>

<div align="center">Residuals $i\sqrt{n}\ \mu$ when the data from table 1 are explained by the model 13.</div>

		Positions				
		P_1	P_2	P_3	P_4	P_5
	H_1	-9.6	-0.7	-1.4	$+3.3$	$+8.4$
Heights	H_2	$+3.6$	$+17.7$	-8.6	-8.3	-4.4
	H_3	-4.2	-2.9	$+2.2$	$+1.1$	$+3.8$

component in the model (13). With respect to this term statistical and technical arguments contradict one another and it would be best to carry out further experiments before reaching a decision.

8. *The function of the analysis of variance*

As pointed out earlier sums of squares, degrees of freedom, and mean squares are not the most suitable form for representing the results of a statistical analysis to technical-minded people. Concrete numerical equations describing the observations plus standard deviations for the residuals these equations do not explain, is what they need. For technical purposes the model expressed in numerical form is of greater interest than the analysis of variance. What then is the function of the analysis of variance in the entire analytical procedure.

In textbooks it is usually stated what is *the model* pertaining to a given experimental design. This, we believe, is a statement which may lead to misinterpretation and which therefore requires some further discussion.

If we set out to investigate a situation where several factors are varied we do not know beforehand what will be the correct model to describe our observations. This is exactly what we wish to find out and the result will depend on what effects turn out to be significant. We must therefore distinguish between two kinds of model, which we may suitably term the *statistical* and the *practical* model.

The model of the textbooks belonging to a specific type of experimental design is the statistical model. It is the most complex model with which the design can adequately deal, the analysis giving one mean square for each of the components in the model.

The *practical model* is the simplest model which provides an adequate description of the observations. When we start the experiment we consider a great variety of conceivable practical models. We must then construct a statistical model which includes all the conceivable practical models and we must design our experiment accordingly.

Looked at from this point of view the analysis of variance serves two purposes.

The main function of an analysis of variance is to decide between a variety of conceivable practical models. Each mean square in an analysis of variance corresponds to a definite term in the equation expressing the model; and whenever a mean square is not significant the corresponding term can be cancelled and the model simplified. Thus by means of the analysis of variance we are able to envisage a great variety of different models and to pick from these at a glance the simplest model that will adequately describe our observations. The examples given above provide illustrations of the procedure. Many others may be taken from the literature.

The second function of the analysis of variance is to provide an estimate

of the residual fluctuations, not accounted for by the practical model finally adopted.

Combining these various remarks we may now set up a few simple rules for using the analysis of variance technique in industry.

A. *Plot the data*

By looking at a plot we can often rule out some models as evidently unsuitable and thereby simplify the analysis. Also in industrial experiments clerical errors or gross errors of observation are not uncommon. A graphical presentation may help to spot and remove outliers, which otherwise may seriously warp our conclusions. The experiment in table 7 and fig. 3 provides an example.

B. *Make up a set of conceivable models*

This should be done on the basis of our technical knowledge and the graphical presentation of the observations.

C. *Decide between the conceivable models by means of an analysis of variance and find the residual variance.*

D. *Express the model chosen in the form of concrete numerical equations.*

In this form the result of the analysis should be sent back to the technicians in the factory.

9. *Levels of significance*

The procedure just outlined may be the subject of statistical criticism. In fact if we use the analysis of variance to test some null-hypothesis it is not permissible to adjust the model to be tested to the observations; the model must be fixed beforehand.

This is of course true. Models adjusted to the observations will generally have a better chance of fitting than predetermined models; that is by the adjustment the probability of obtaining significance will be increased.

It is not likely, however, that this increase will be such as to invalidate the conclusions drawn. On the whole critical testing of significance is of secondary interest. Usually the analysis of variance mainly serves to provide a general survey as to what effects are pronounced but a final decision which of the significant effects are to be incorporated in the model will often be made rather by technological than by statistical argument.

Hence if by the procedure of the previous section the levels of significance are somewhat violated, this should not be a matter of serious concern.

10. *Normality; computing the residual*

This remark also furnishes a reply as to conditions of normality. Tests of significance usually assume that the random fluctuations are normally distributed. But considerable deviations from normality will cause only comparatively slight changes in the significance levels; so we need not be seriously concerned on that score.

Deviations due to reading or clerical errors, or to outlying observations caused by abnormal experimental conditions are more serious. We have already pointed to the importance of plotting the data before embarking on a numerical analysis in order to detect outliers.

An alternative method for checking is by computing the elements of the residual as we have done in tables 4, 12, 14, and 16. Sometimes the distribution of positive and negative signs among the residuals indicate that the model is not quite adequate; tables 12 and 16 furnish examples. Also the values of the residuals may assist in locating outliers; for example the residual of 17.7 in table 16 is a bit high; it is more than twice the estimated standard deviation of a single observation ($s = 8.4$; $\nu = 10$). This, however, is not a correct test. The value 17.7 is a residual with respect to the adjusted equation 13 and the standard deviation of such residuals is on the average smaller than that of the observations themselves. Besides the 17.7 residual is included in s, and if it is abnormally high it will have produced an abnormal increase in this estimate. It would not be easy to give a precise test, but we may certainly consider the residual 17.7 as suspect.

If the residuals are sufficiently numerous they can be plotted in a histogram which should approximately have the shape of a normal distribution. In many instances we have found the computation of the residuals not too laborious and really helpful in checking and understanding the underlying analysis.

11. *Testing for interaction in a two-way classification with unknown levels in both directions*

As pointed out in §5 we can separate interaction from random fluctuations by replication. We will now briefly discuss an alternative method which has been proposed by Tukey[6] and Ward and Dick[7].

For known levels in both directions the leading interaction term is given by the product $X_i Y_i$ as discussed in §7. If we do not know the levels we can still test for interaction by using the leading differences $(Z_i. - Z..)$ and $(Z._i - Z..)$ computed from the observations for rows and columns as estimates for X_i and Y_i. In keeping with the methods of §7 we then obtain a regression coefficient.

$$b_{11} = \frac{\sum\limits_{ij} (Z_{i.} - Z..)(Z_{.j} - Z..)Z_{ij}}{\sum\limits_{ij} \{(Z_{i.} - Z..)(Z_{.j} - Z..)\}^2} \qquad (14)$$

and a corresponding sum of squares

$$V_{11} = \frac{\{\sum\limits_{ij} (Z_{i.} - Z..)(Z_{.j} - Z..)Z_{ij}\}^2}{\sum\limits_{ij} \{(Z_{i.} - Z..)(Z_{.j} - Z..)\}^2}. \qquad (15)$$

This is Tukey's method. A more satisfactory procedure would be to adopt the model

$$Z = g + a_i + b_j + ca_ib_j + \epsilon_{ij}, \qquad (16)$$

as proposed by Ward and Dick[7]. This model can not be adjusted in a simple way, but Ward and Dick have developed an iterative procedure, the first stage of which is equivalent to Tukey's method. Since Ward and Dick have not provided numerical examples of their method it seems appropriate to do so here. The formulae are given in the Appendix.

First let us reconsider the data of table 10 interchanging rows and columns in random order so that the systematic variations in row and column averages are lost (table 17). To fix our thoughts let us imagine that we have five batches of the same type of resistors which have been subjected to a heat treatment on five different occasions. It is quite conceivable that variations in humidity during production interact with humidity variations during the heat treatment; Tukey's or Ward and Dick's procedure might reveal such interaction.

TABLE 17

Data of table 10 with rows and columns interchanged in random order and ascribed to imaginary factors.

Batches of resistors		1	2	3	4	5	
			Per mil changes in resistance				$Z_{i.}$
Heat Treatment	1	1.9	2.2	2.1	1.7	4.1	2.4
	2	2.0	3.6	3.4	3.9	3.1	3.2
	3	3.2	3.4	2.8	4.2	1.9	3.1
	4	2.1	4.3	2.6	6.1	3.4	3.7
	5	2.8	2.5	2.6	2.1	3.0	2.6
	$Z_{.j}$	2.4	3.2	2.7	3.6	3.1	3.0

In table 18 the mean squares resulting from successive stages of Ward and Dick's analysis have been entered together with those resulting from our previous analyses; table 19 contains the successive values of the constants a_i, b_j and c.

If we introduce linear components as in table 13 we find significant effects for rows, columns, and interaction. If we analyze into rows and columns, the significance of row and column differences disappears because the greater part of a sum of squares which was first concentrated into one degree of freedom is now spread out over 4 degrees of freedom.

TABLE 18

Ward and Dick's analysis applied to the data of table 17; mean squares in (per mil)^2

Analysis	Columns = Batches		Rows = Treatments		Interaction		Residual	
	M.S.	D.F.	M.S.	D.F.	M.S.	D.F.	M.S.	D.F.
With linear components as in table 13	4.20	1	4.50	1	7.13	1	0.40	21
Ordinary analysis into rows and columns	1.07	4	1.32	4	—	—	0.93	16
Ward and Dick's analysis 1st Stage = Tukey's method	1.07	4	1.32	4	6.15	1	0.58	15
2nd Stage	1.07	4	1.32	4	12.49	1	0.16	15
3rd Stage	0.89	4	1.31	4	9.61	1	0.41	15
4th Stage	0.87	4	1.31	4	10.51	1	0.35	15
5th Stage	0.88	4	1.32	4	10.58	1	0.34	15

But Tukey's analysis still reveals the interaction though somewhat less pronounced than before.

The significance of this interaction is greatly enhanced in the subsequent stages of Ward and Dick's analysis.

In the second stage the mean square rises to 12.49 but is reduced again in the later stages. This may seem surprising but it should be borne in mind that these mean squares are computed from formulae which are correct only when the constants in the model have been fully adjusted, but do not hold for the intermediate stages of approximation. Hence we must judge by the final results.

It will be seen from tables 18 and 19 that after 4 stages we reach a reasonable constancy; the changes from the 4th to the 5th stage are relatively unimportant.

TABLE 19

Adjusted constants when Ward and Dick's analysis is applied to the data of table 17.

		Stage in the analysis				
		1st	2nd	3rd	4th	5th
	$\hat{a}_1$	−0.6	−0.460	−0.538	−0.536	−0.550 per mil
Rows =	$\hat{a}_2$	+0.2	+0.162	+0.137	+0.142	+0.138
Heat treatments	$\hat{a}_3$	+0.1	−0.009	+0.129	+0.117	+0.140
	$\hat{a}_4$	+0.7	+0.795	+0.720	+0.726	+0.716
	$\hat{a}_5$	−0.4	−0.488	−0.448	−0.449	−0.443
	$\hat{b}_1$	−0.6	−0.489	−0.407	−0.380	−0.383
Columns =	$\hat{b}_2$	+0.2	+0.230	+0.180	+0.185	+0.187
Batches	$\hat{b}_3$	−0.3	−0.247	−0.210	−0.205	−0.206
	$\hat{b}_4$	+0.6	+0.792	+0.643	+0.659	+0.666
	$\hat{b}_5$	+0.1	−0.286	−0.206	−0.260	−0.264
Interaction	$\hat{c}$	+0.260	+0.370	+0.360	+0.380	+0.378

From a technical point of view it is to be observed that when in a practical case with unknown levels in both directions we find such a pronounced interaction as in the last row of table 18 but no significant effects between rows or columns, this may be taken as a strong indication that definite factors have been operative in the rows and columns, and that we have failed to find row and column effects only because these factors have not been taken into account and their effect has been spread out over too many degrees of freedom.

A second instructive example is provided by the observations on layer thicknesses of Al_2O_3 coatings recorded in table 1; the final results after a five-stage iteration are given in table 20.

In the complete analysis by orthogonal polynomials (table 15) we found $\xi_1\eta_2$ and $\xi_2\eta_1$ to be the two significant interaction components with a total sum of squares of 949. Of this only the amount 512 is revealed by the Ward and Dick analysis which does consequently not detect interaction effects as effectively as does the analysis of §7. To study this point somewhat more in detail the analysis by orthogonal polynomials was also applied to the residuals with respect to Ward and Dick's model, computed with the aid of the adjusted constants of table 20A. This gave the results recorded in table 21.

TABLE 20

Final results obtained by applying a 5-stage Ward and Dick analysis to the data of table 1

A. Adjusted constants, in microns

Heights		Positions		Interaction
$_5\hat{a}_1$	− 6.46	$_5\hat{b}_1$	−8.38	
$_5\hat{a}_2$	− 6.45	$_5\hat{b}_2$	+2.82	
$_5\hat{a}_3$	+13.00	$_5\hat{b}_3$	+6.89	$_5\hat{c} = 0.1124$
		$_5\hat{b}_4$	+3.44	
		$_5\hat{b}_5$	−4.77	

B. Mean squares

Source	S.S.	D.F.	M.S.	
Heights	1268	2	634	(microns)^2
Position	480	4	120	
Interaction	512	1	512	
Residual	798	7	114	

TABLE 21

Sum of squares corresponding to the $\xi_1\eta_2$ and $\xi_2\eta_1$ interaction components in the original data of table 1 and in their Ward and Dick residuals

Component	Sum of squares	
	Original data	Ward and Dick residual
$\xi_1\eta_2$	603	45
$\xi_2\eta_1$	346	298

We see that the Ward and Dick analysis has chiefly removed the $\xi_1\eta_2$ interaction and not the $\xi_2\eta_1$ interaction.

In this connection it is to be noted that in table 15 and the model (13) of the "pure" terms, containing X or Y alone, only $\xi_1\eta_0$, $\xi_2\eta_0$, and $\xi_0\eta_2$ occurred, while the linear term in Y, $\xi_0\eta_1$, is missing; the sum of squares for this term was very small, $11\mu^2$ only. Table 21 therefore

suggests that the Ward and Dick analysis reveals interactions of the type $\xi_k \eta_l$ only if the corresponding pure components $\xi_0 \eta_l$ and $\xi_k \eta_0$ are both pronouncedly present.

This is reasonable enough, though our attempts to prove it have failed. If we represent the observations by their orthogonal components, $Z = \sum\sum r_{ij} \, \xi_i \eta_j$, it is fairly easy to show that interactions $\xi_k \eta_l (k > 0, l > 0)$ are contained in the first stage estimates $_1\hat{a}_i$ and $_1\hat{b}_j$, and contribute to the interaction constant $_1\hat{c}$ only if both the pure components $\xi_k \eta_0$ and $\xi_0 \eta_l$ are contained in Z. What happens at later stages becomes complex, however, and is not mathematically very tractable.

It should also be observed that Tukey's analysis yielded a mean square for interaction of $190\mu^2$ against $512\mu^2$ for Ward and Dick's. The latter method of analysis is therefore in this instance much more effective.

Of course the applications we have made above are somewhat artificial, because we have applied an analysis for unknown levels to cases where in reality the levels were known. How often cases may occur where Ward and Dick's analysis will reveal interaction effects which could not be discovered otherwise, it is difficult to say. But there can be little doubt that this method is an interesting addition to our arsenal of statistical techniques, which it is well worth trying out in practice.

12. *Conclusions*

This paper does not contain any material that is essentially new. Examples of the various types of analysis of §3, and 7 have occasionally been published, but they have never been brought together and discussed from a unifying point of view.

Even a two-way classification may give rise to a great variety of models and analyses. All the situations described occur in industrial applications and to deal with industrial problems we must have the whole gamma of techniques discussed above readily available. In this respect the treatment of the two-way classification in existing textbooks is not quite satisfactory.

It will be clear that if we pass on to three-way or four-way classifications the variety of cases increases tremendously, so much so that a systematization of conveivable models and methods of analysis seems almost impossible. The situations encountered in practice are never exactly the same, and in each case we have to think out afresh what models and what kind of analysis can best be applied. Sometimes it even requires some trial-and-error analysis before a satisfactory solution is obtained.

But in order to solve successfully these complex problems the implications of a two-way classification should first of all be fully understood. The present paper was written with this point in mind.

I wish to conclude by expressing my sincere thanks to Mr. A. M. van Beek for his valuable assistance in carrying out the numerical analyses.

<div align="right">Eindhoven, December 3rd, 1954</div>

LITERATURE

1. A. Palazzi, La progettazione fattoriale delle ricerche industriali; exempi di applicazione. *La Metallurgica Italiana 6*, 215–220, 1952.
2. O. L. Davies a.o., *Design and analysis of industrial experiments*, Oliver and Boyd, Edinburgh, 1954.
3. J. W. Tukey, Interaction in a row-by-column design, Memorandum Report 18, Statistical Research Group, Princeton University, 1949.
4. D. B. DeLury, *Values and integrals of the orthogonal polynomials up to $n = 26$*, University of Toronto Press, 1950.
5. R. A. Fisher and F. Yates, *Statistical tables for biological, agricultural and medical research*, Oliver and Boyd, Edinburgh, 1948.
6. J. W. Tukey, One degree of freedom for non-additivity, *Biometrics 5*, 232–242, 1949.
7. G. C. Ward and I. D. Dick, Non-additivity in randomized block designs and balanced incomplete block designs, *New Zealand Journal of Science and Technology 33*, 430–435, 1952.

APPENDIX: The iterative procedure of Ward and Dick

Model:
$$Z = g + a_i + b_j + ca_i b_j + \epsilon_{ij} \tag{17}$$

Symbols:

Z_{ij} = the observations;

$i = 1, 2, \cdots , m$ indicating rows,

$j = 1, 2, \cdots , n$ indicating columns

m = number of rows,

n = number of columns.

$S_{i.} = \sum_j Z_{ij}$ = Sum of the observations in the i-th row,

$S_{.j} = \sum_i Z_{ij}$ = Sum of the observations in the j-th column,

$S_{..} = \sum_i \sum_j Z_{ij}$ = Sum of all observations,

$\left. \begin{array}{l} {}_k\hat{a}_i \\ {}_k\hat{b}_j \\ {}_k\hat{c} \end{array} \right\}$ estimates of a_i , b_j and c obtained after k iterations

Formulae

$$_{(k+1)}\hat{a}_i = \frac{S_{i.} - S_{..}/m + {}_k\hat{c} \sum_j {}_k\hat{b}_j Z_{ij} - ({}_k\hat{c}/m) \sum_j {}_k\hat{b}_j S_{.j}}{n + ({}_k\hat{c}^2/m) \sum_j {}_k\hat{b}_j S_{.j}} , \qquad (18)$$

$$_{(k+1)}\hat{b}_j = \frac{S_{.j} - S_{..}/n + {}_k\hat{c} \sum_i {}_k\hat{a}_i Z_{ij} - ({}_k\hat{c}/n) \sum_i {}_k\hat{a}_i S_{i.}}{m + ({}_k\hat{c}^2/n) \sum_i {}_k\hat{a}_i S_{i.}} , \qquad (19)$$

$$_{(k+1)}\hat{c} = \frac{mn \sum_i \sum_j {}_{(k+1)}\hat{a}_i {}_{(k+1)}\hat{b}_j Z_{ij}}{\sum_i {}_{(k+1)}\hat{a}_i S_{i.} \sum_j {}_{(k+1)}\hat{b}_j S_{.j}}. \qquad (20)$$

The estimate of g is

$$\hat{g} = S_{..}/mn. \qquad (21)$$

The iterative process is started with

$$_0\hat{c} = 0. \qquad (22)$$

We then compute $_1\hat{a}_i$, $_1\hat{b}_j$, from these $_1\hat{c}$ and so on. A check on the computations at each stage is provided by the equations

$$\sum_i {}_k\hat{a}_i = 0, \qquad (23)$$

$$\sum_j {}_k\hat{b}_j = 0. \qquad (24)$$

The reduction in the sum of squares is given by

$$R = \hat{g}S_{..} + \sum_i \hat{a}_i S_{i.} + \sum_j \hat{b}_j S_{.j} + \hat{c} \sum_{ij} \hat{a}_i \hat{b}_j Z_{ij} \qquad (25)$$

As pointed out in §11 this formula only holds exactly for the final estimates $\hat{a}_i$, $\hat{b}_j$, $\hat{c}$, but is not correct for intermediate estimates $_k\hat{a}_i$, $_k\hat{b}_j$, $_k\hat{c}$.

Example

As an example we consider the data of table 1. First of all we find

$$\hat{g} = 137, \qquad (26)$$

and all further computations may be simplified by subtracting a constant, for instance 100, from each datum. The basic data then take the form of table 22.

TABLE 22

Basic data for computation according to Ward and Dick's model.

$\rightarrow$ j $\downarrow$ i	Z_{ij}					$S_{i.}$	$S_{i.} - S_{..}/m$
	25	30	28	34	43	160	-25
	26	50	27	24	18	145	-40
	30	55	68	59	38	250	$+65$
$S_{.j}$	81	135	123	117	99	555	
$S_{.j} - S_{..}/n =$	-30	$+24$	$+12$	$+6$	-12	$= S_{..}$	

$$m = 3; \quad n = 5$$

The successive computations may then conveniently be arranged as in table 23, which may be continued towards the right for subsequent stages.

TABLE 23

Computing scheme for the Ward and Dick analysis.

$i =$	$S_{i.} - S_{..}/m$	$_1\hat{a}_i$	$\sum_j {}_1\hat{b}_j Z_{ij}$	$\sum_i \sum_j {}_1\hat{a}_{i1}\hat{b}_j Z_{ij}$	$_2\hat{a}_i$
1	-25	-5	-2		-6.49
2	-40	-8	224	3132	-6.72
3	$+65$	$+13$	378		$+13.21$

$$\sum_j {}_1\hat{b}_j S_{.j} = 600 \qquad {}_1\hat{c} = \frac{3.5.3132}{600.1290} =$$

$j =$	$S_{.j} - S_{..}/n$	$_1\hat{b}_j$	$\sum_i {}_1\hat{a}_i Z_{ij}$	$= 0.0607$	$_2\hat{b}_j$
1	-30	-10	57		-10.68
2	$+24$	$+8$	165		$+4.65$
3	$+12$	$+4$	528	3132	$+7.19$
4	$+6$	$+2$	405		$+3.78$
5	-12	-4	135		-4.93

$$\sum_i {}_1\hat{a}_i S_{i.} = 1290$$

$m = 3, n = 5$.

$\sum_i \sum_i {}_1\hat{a}_i {}_1\hat{b}_j Z_{ij}$ is computed as $\sum_i {}_1\hat{a}_i \sum_j {}_1\hat{b}_j Z_{ij}$ and as $\sum_j {}_1\hat{b}_j \sum_i {}_1\hat{a}_i Z_{ij}$ to check the arithmetic.

THE EXPLORATION AND EXPLOITATION OF RESPONSE SURFACES:

AN EXAMPLE OF THE LINK BETWEEN THE FITTED SURFACE AND THE BASIC MECHANISM OF THE SYSTEM

G. E. P. Box and P. V. Youle

Imperial Chemical Industries Limited,
Dyestuffs Division Headquarters,
Blackley, Manchester, England

This is a sequel to an article which recently appeared in this journal [1] and had the same general title. The previous article described a number of applications of newly developed techniques [2] for the study of response surfaces. The present article shows how study of the form of the empirical surface can throw important light on the basic mechanism operating and can thus make possible developments in the *fundamental theory* of a process. This idea is illustrated in some detail with an example previously discussed only from the empirical standpoint. A theoretical surface, based on reaction kinetics is now derived, rate constants are estimated from the data and the theoretical surface is compared with the empirical surface previously obtained. It is then shown how the canonical variables of the empirical surface can relate to the basic physical laws controlling the system. In this connection the problem of suitable choice of metrics for the variables is discussed. In a final section some general remarks on the process of scientific investigation are appended.

I. INTRODUCTION

A response surface is a graphical representation of a relationship

$$\eta = \phi(x_1, x_2, \cdots, x_k)$$

between some response such as yield, whose level is denoted by η, and a number of quantitative variables (or factors), such as temperature, time and concentration, whose levels are denoted by $x_1, x_2, \cdots, x_k$.

The feature of the surface of greatest interest is often the value or values of the variables $x_1, x_2, \cdots x_k$ for which η is a maximum.

In the previous paper it was emphasised that the study of numerous examples had indicated that sharply defined point maxima appeared to be something of a rarity. The typical situation was that in which the response was found to be insensitive in the region of the maximum

287

to certain joint changes in the levels of variables, indicating the existence of 'factor dependence'. An extreme form of this phenomenon occurred where there was a line, plane, or space of near-maxima rather than a single point maximum. Such a response surface was said to contain a 'stationary ridge system'. A second type of surface of common occurrence contained a rising ridge. It was suggested that the nature of a ridge system could indicate the physical laws which underlay the process studied.

The method recommended for exploring a response surface consisted first of performing a simple pattern of experiments designed to detect, in the initial region explored, any general sloping tendency of the surface. If such a tendency was found, further experiments were performed in the indicated direction of increasing response. Either initially, or after one or two cycles of this 'steepest ascent' procedure had brought the experimenter to a region of higher response, it was usually found that no sloping tendency could be detected and exploited. The region so attained was then examined by performing a slightly more elaborate pattern of experiments and fitting a suitable function which enabled curvature in the surface and dependence between the variables to be taken account of.

In the absence of prior knowledge concerning the form of the response function, a local representation could be obtained by fitting a polynomial in x_1, x_2, $\cdots$, x_k, in which all terms up to a given order d were included. This was of course equivalent to supposing that the true function could be locally represented to a sufficient approximation by its Taylor series ignoring terms of order higher than d.

In the majority of applications, where the object was not so much to graduate the response surface accurately but rather to determine approximately its general characteristics in the optimum region, an equation of only second degree has usually been adequate.* Reduction of this fitted second degree equation to canonical form has allowed the nature of the fitted surface to be readily appreciated and has indicated in what regions further experiments were necessary.

It has been found that:

(1) This approach has made it possible to comprehend features of the surface which could be exploited to attain further gain when possibility of improvement by simpler means had been exhausted.

(2) By considering the features of the surface for the principal response such as yield, or cost, in relation to the features of the surface for 'auxiliary' responses such as purity, it has been possible to discover

---

*Where more accurate *graduation* was required (as for example in the work on pulse columns performed for the Atomic Energy Commission) an equation of third degree has been used [4] [5].

conditions which were 'best' in the practical sense of bringing all the responses to 'most satisfactory' compromise levels.

(3) Consideration of the shape of a fitted response surface has suggested new theories of behaviour of the system.

It is this last aspect which we shall here discuss further.

In the analysis of the fitted second-degree equation the existence of a ridge is indicated by one or more of the coefficients in the canonical form of the equation being small in comparison with the others. Where these small coefficients are negligible it is implied that the system in k variables can be more economically described in terms of less than k canonical or 'compound' variables. It appears that these compound variables can have greater significance than a purely representational one. In fact they can indicate the fundamental mechanism of the system. To make this clear we first consider a simple hypothetical example.

Suppose that the effect on yield of the concentrations c_1 and c_2 of two reactants were being studied and that previous experimentation had suggested that we should now explore the ranges of concentration: $c_1 = 50\text{--}60$ grams per litre and $c_2 = 30\text{--}40$ grams per litre which were expected to be near their optimum values. It is usually simplest in such examples to work with coded values of the variables and we will suppose that 'standardised variables' were chosen as follows:

$$x_1 = (c_1 - 55)/5 \qquad x_2 = (c_2 - 35)/5$$

so that the region explored with suitably placed experiments was defined by

$$-1 < x_1 < +1, \qquad -1 < x_2 < +1$$

Suppose finally that the fitted second degree equation was

$$Y = 78.56 + 0.50x_1 - 0.21x_2 - 2.31x_1^2 - 2.15x_2^2 + 4.08x_1x_2$$

and that the errors of estimate of the coefficients were sufficiently small so that the equation was as a whole meaningful.

Now this, like any other second degree equation, can be written in canonical form. (That is by changing the origin and rotating the co-ordinate axes we can write it in form containing only quadratic terms). In the present case the equation, written in this way becomes

$$Y = 78.63 - 4.27X_1^2 - 0.19X_2^2$$

where

$$X_1 = 0.72x_1 - 0.69x_2 - 0.06$$

$$X_2 = 0.69x_1 + 0.72x_2 - 0.53$$

(1)

and these last two equations define the positions and directions of the new coordinate axes.

The centre of the system (that is the point $X_1 = 0$, $X_2 = 0$) has co-ordinates $x_1 = 0.41$, $x_2 = 0.34$. Thus the axes of the system defined by the lines $X_1 = 0$ and $X_2 = 0$ pass close to the original origin and through the region in which the experiments have been performed. A discussion of the surface in terms of these axes is relevant therefore.

We notice that very nearly the equation is:

$$Y = 78.6 - 4.27(0.7)^2(x_1 - x_2 - 0.1)^2 - 0.19(0.7)^2(x_1 + x_2 - 0.8)^2$$

or in terms of the original units

$$Y = 78.6 - 0.084(c_1 - c_2 - 20.5)^2 - 0.004(c_1 + c_2 - 94.0)^2$$

Thus the canonical variables correspond very nearly, to the *difference* of the concentrations and the *sum* of the concentrations, the coefficient of the difference being much larger than that of the sum. Now remembering that our estimates of the coefficients are subject to error and also that the form of equation is probably not entirely adequate, it would seem that the data might be explained on the hypothesis that yield depended *only* on the difference between the concentrations and not at all on their 'overall' level, the best yield being attained whenever c_1 was about 20 grams per litre greater than c_2. This hypothesis could be readily checked over wider ranges of the variables by further experiment.

Assuming this hypothesis was shown to be substantially correct, attention which had so far been focussed on the mathematical analysis would be shifted to physico-chemical theory. The experimenter would ask himself "What mechanism could produce the phenomenon of yield being dependent on this concentration difference?" If he could answer that question further experiments would be devised to test his theory. Such a theory by contributing to a basic understanding of the mechanism of the reaction could, for example, lead to new methods of overcoming yield-limiting factors either by modification of the physical conditions or by the introduction of other reactants into the system. The fitted equation may thus provide not merely an empirical representation of the surface near the maximum (useful though this is) but also a valuable indication of how the system works.

Now as a consequence of the form of our fitted equation the canonical variables X_1 and X_2 are necessarily expressed as *linear functions* of the quantities x_1 and x_2 but it is obvious that usually an underlying 'compound variable' will be some less simple function. For example it will

frequently happen that the level of yield will depend on the *ratio* of two concentrations rather than their difference and we shall see later that in real examples more complicated relationships occur. The difficulties which this presents are not as serious as they first appear.

Let us consider the particular example of the yield depending on the ratio of the concentrations of two reactants, there being a certain optimum ratio. A second degree equation fitted to the *logarithms* of c_1 and c_2 would give an equation similar to that obtained before but with a dominant canonical variable $(\log c_1 - \log c_2)$ instead of $(c_1 - c_2)$ and the yield surface plotted in terms of $\log c_1$ and $\log c_2$ would contain a stationary ridge system.

Now in practice we should usually be attempting to represent the relationship over ranges of concentration which were fairly small compared with the overall magnitudes of the concentrations. The appearance of the surface would then not be very different whether it was plotted in terms of c_1 and c_2 or in terms of $\log c_1$ and $\log c_2$ and a ridge system which was represented by equations like (1) would still be found even though the second degree equation were fitted to c_1 and c_2 rather than to $\log c_1$ and $\log c_2$. That this is generally true for other types of functions can readily be seen.

Suppose Y depends only on $X = f(c_1, c_2)$ and $f(c_1, c_2)$ can be represented locally reasonably well by the first order terms of its Taylor series

$$X = f_0(c_1, c_2) + (\partial f/\partial c_1)_0 (c_1 - c_{10}) + (\partial f/\partial c_2)_0 (c_2 - c_{20}) \quad (2)$$

where the subscript zero denotes that the value at the point c_{10}, c_{20} is taken. Now this equation is of the linear form

$$X = ac_1 + bc_2 + d \qquad (3)$$

so that if, in the region of the optimum, Y, could be approximated by a quadratic function of X, we should have

$$Y = Y_0 + B(X - X_0)^2$$

We see therefore that while we should expect that our procedure for *detecting* local factor dependence would have fairly wide applicability, the question of what was the relevent form for the compound variable would usually be a matter for further speculation and experiment.

Returning to the case where yield is dependent on the level of c_1/c_2 equation (2) gives for points in the neighbourhood of c_{10}, c_{20}

$$\frac{c_1}{c_2} \simeq \frac{c_{10}}{c_{20}}\left[1 + \frac{c_1}{c_{10}} - \frac{c_2}{c_{20}}\right] \qquad (4)$$

Here therefore the experimenter working with the untransformed variables c_1 and c_2 would find a dominant canonical variable $X = ac_1 + bc_2 + d$ in which a/b was roughly equal to c_{20}/c_{10} the ratio of *the average concentrations*. This would indicate that proportional changes in the concentrations would leave X and hence the yield Y unchanged and would lead to recalculation of the equation in terms of $\log c_1$ and $\log c_2$ when a closer fit would probably be obtained.

Usually where the first analysis indicates some ridge system we must rely on possible theories of the system to indicate the correct function to employ. These theories must of course be checked by further experimentation.

The simple numerical illustration quoted above is hypothetical but serves to introduce the following genuine but more complicated example.

2. THE EMPIRICAL STUDY

An experimental study of the system concerned has been described in reference [1] (first example) and some of the detailed calculations will be found in [3].* It was desired to maximise the yield of one of the products of a chemical reaction. To do this the yields of this and other products of the reaction were experimentally determined under various conditions of temperature (T), concentration of one of the reactants (c), and time of reaction (t). The conditions T, c and t were measured as degrees centigrade, % concentration and 'hours on temperature' respectively.

The results are shown in columns 1, 2, 3, 4, 11, 12, 13 and 14 of Table 1. The quantity $\hat{\eta}_2$ is the estimated fraction of unchanged starting material, $\hat{\eta}_3$ the estimated fraction converted to the desired product and $\hat{\eta}_5$ the estimated fraction occurring as an unwanted by-product. The fractions are called the 'yields' and are sometimes quoted as percentages. The circumflex accent will be used throughout this paper to indicate *observed* or *estimated* quantities, the 'true' values will be unaccented.

For convenience the levels of the variables are coded in columns (5), (6) and (7) of Table 1 as follows:

$$x_1 = (T - 167)/5, \qquad x_2 = (c - 27.5)/2.5, \qquad x_3 = (t - 6.5)/1.5$$

The coded values for the first eight experiments are then all at the levels $+1$ and -1 forming a 2^3 factorial design. When this is augmented with experiments 9–15 a 'central composite' experimental design [2]

---

*The 'natural units' given in [3] differ slightly from those quoted in reference [1]. The yields given in Table 1 of this paper differ slightly from those in [3] due to refinements previously ignored.

TABLE I. EXPERIMENTAL DATA

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	Levels of Variables			First Coding			Second Coding			Observed Yields						
Expt.	$T(^\circ C)$	$c(\%)$	$t(hr.)$	x_1	x_2	x_3	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_3$	η_2	η_3	η_5	Total	$\hat{z}_t$	$\mathfrak{q}$	$x = (T+273)^{-1}$
1	162	25	5	−1	−1	−1	−1	−1	−1	0.415	0.459	0.112	0.986	0.739	−6.29	$10^{-7} \times 22989$
2	162	25	8	−1	−1	1	−1	−1	1	0.338	0.533	0.112	0.983	0.714	−6.64	22989
3	162	30	5	−1	1	−1	−1	1	−1	0.277	0.575	0.127	0.979	0.683	−6.23	22989
4	162	30	8	−1	1	1	−1	1	1	0.217	0.588	0.160	0.965	0.632	−6.49	22989
5	172	25	5	1	−1	−1	1	−1	−1	0.199	0.606	0.162	0.967	0.627	−5.80	22472
6	172	25	8	1	−1	1	1	−1	1	0.150	0.580	0.226	0.956	0.559	−6.04	22472
7	172	30	5	1	1	−1	1	1	−1	0.122	0.586	0.245	0.953	0.536	−5.70	22472
8	172	30	8	1	1	1	1	1	1	0.043	0.524	0.380	0.947	0.413	−5.79	22472
9	167	27.5	6.5	0	0	0	0.01	0.05	0.12	0.193	0.569	0.213	0.975	0.595	−6.05	22727
10	177	27.5	6.5	2	0	0	1.97	0.05	0.12	0.064	0.554	0.308	0.926	0.391	−5.42	22222
11	157	27.5	6.5	−2	0	0	−2.03	0.05	0.12	0.376	0.469	0.147	0.992	0.707	−6.51	23256
12	167	32.5	6.5	0	2	0	0.01	1.92	0.12	0.180	0.575	0.222	0.977	0.586	−6.21	22727
13	167	22.5	6.5	0	−2	0	0.01	−2.19	0.12	0.263	0.550	0.183	0.996	0.651	−6.05	22727
14	167	27.5	9.5	0	0	2	0.01	0.05	1.73	0.099	0.589	0.280	0.968	0.515	−6.18	22727
15	167	27.5	3.5	0	0	−2	0.01	0.05	−2.52	0.250	0.503	0.221	0.974	0.605	−5.47	22727
16	177	20	6.5	2	−3	0	1.97	−3.54	0.12	0.141	0.611	0.230	0.982	0.560	−5.60	22222
17	177	20	6.5	2	−3	0	1.97	−3.54	0.12	0.152	0.629	0.207	0.988	0.596	−5.72	22222
18	160	34	7.5	−1.4	2.6	0.7	−1.41	2.36	0.72	0.159	0.600	0.221	0.980	0.583	−6.39	23095
19	160	34	7.5	−1.4	2.6	0.7	−1.41	2.36	0.72	0.196	0.606	0.193	0.995	0.624	−6.54	23095

[1] is formed suitable for fitting an equation of second degree to any observed response.

A second degree equation fitted to the yields of desired product $\hat{\eta}_3$ for experiments 1–15 indicated a possible planar ridge system. Further experiments 16–19 were carried out therefore on the estimated maximum plane. In spite of the great differences in the actual conditions employed (see table 1) these gave yields close to the maximum value of about 60% in accordance with prediction. These observations were now included in the calculation and the best fitting second degree equation using all 19 observations was then:

$$Y = 58.78 + 1.90x_1 + 0.97x_2 + 1.06x_3 - 1.88x_1^2 - 0.69x_3^2$$
$$- 0.95x_2^2 - 2.71x_1x_2 - 2.17x_1x_3 - 1.24x_2x_3 \qquad (5)$$

where Y denotes the % yield predicted by the equation.

From Table 2 it will be seen that the sum of squares due to regression accounted for 92.6% of the total variation after elimination of the mean. From the residual sum of squares an estimate $\sigma = 1.81$ of the experimental error standard deviation was obtained (this may be biased upwards due to some inadequacy of the assumed form of the equation). Using this estimate for σ the standard errors of the coefficients in equation (5) could be calculated. For linear, quadratic and interaction terms these were all between 0.3 and 0.5. It appeared therefore that equation (5) was reasonably well determined. It was found to have the canonical form

$$Y - 59.15 = -3.40X_1^2 - 0.32X_2^2 + 0.20X_3^2 \qquad (6)$$

where
$$X_1 = 0.751x_1 + 0.479x_2 + 0.455x_3 - 0.349 \qquad (7)$$

$$X_2 = 0.308x_1 + 0.356x_2 - 0.882x_3 + 0.013 \qquad (7a)$$

$$X_3 = 0.584x_1 - 0.803x_2 - 0.120x_3 + 0.485 \qquad (7b)$$

The centre of the system (that is the point $X_1 = 0$, $X_2 = 0$, $X_3 = 0$) had coordination $x_1 = -0.03$, $x_2 = 0.55$, $x_3 = 0.23$. Thus the axes of the system passed through the region in which the experiments had been performed and a discussion in terms of these axes was therefore immediately relevant.*

---

*In the case of ridges, especially rising ridges, the 'centre' of the canonical system may be found almost anywhere on the line or plane of the crest of the ridge. When this centre is remote from the region of experiments, we cannot, of course, draw any conclusions about the nature of the surface at this remote point. However the preliminary use of the steepest ascent procedure will have ensured that the experimenter has already been brought to a point which is close to the ridge. The canonical equation can therefore be rewritten in terms of a new origin close to the original origin and on the line or plane of the ridge. From this form of the equation the principal features of the response surface in the *immediate region* of the experiments may be readily comprehended (see for example reference [1] pp. 37 and 53 and reference [3] p. 531.

Now the canonical variables have the same scale as the original variables. That is to say in solid models (like those in Figures 3 and 4) in which unit change in x_1, x_2, or x_3 is represented by the same distance, unit change in X_1, X_2 and X_3 is also represented by this same distance. Consequently the relative magnitudes of the coefficients in the canonical equation indicate the relative *importance* of the canonical variables in describing the function over the experimental region which has been chosen as appropriate for study. The coefficient (-3.40) of X_1^2 in equation (6) is over 10 times larger than either of the other two coefficients (-0.32 and 0.20). Furthermore the latter are somewhat less in magnitude than their errors of estimation. (Appropriate standard errors for these constants can be shown to be of the same order of magnitude as those of the original quadratic and interaction terms, namely about 0.3 to 0.5.) To express the matter a little differently within the region in which the fitted equation has some relevance which may be roughly defined as $-2 < X_1 < +2$, $-2 < X_2 < +2$, $-2 < X_3 < +2$, the maximum contribution to the % yield Y of the terms in X_1 is about 14% whilst that of each of the terms in X_2 and X_3 is only about 1% which is of the same order of magnitude as the experimental standard deviation. These facts suggest that the system may be described by an equation containing a single canonical variable only.

The refitting *ab initio* of an equation of the form

$$Y - Y(\text{max}) = BX^2$$

containing only a single variable X which is itself linear in x_1, x_2 and x_3 is possible but laborious. An approximation used in [1], [2] and [3] may be obtained simply by ignoring the smaller coefficients in equation (6) when we have

$$Y - 59.15 = -3.40X_1^2$$

with X_1 defined as before (equation (7)).

A closer approximation is obtained by fitting by least squares the expression

$$Y = A + BZ + CZ^2 \tag{8}$$

where Z is the linear aggregate $ax_1 + bx_2 + cx_3$ obtained by omitting the constant term in X_1. In the present example

$$Y = 59.50 + 2.65Z - 3.80Z^2 \tag{9}$$

where $$Z = 0.751x_1 + 0.479x_2 + 0.455x_3 \tag{9a}$$

That is
$$Y - 59.96 = -3.80X_2'^2 \qquad (10)$$

where
$$X_1' = 0.751x_1 + 0.479x_2 + 0.455x_3 - 0.348 \qquad (11)$$

or in terms of the original variables

$$X_1' = 0.150T + 0.191c + 0.303t - 31.969 \qquad (11a)$$

An analysis of variance is shown in table 2.

TABLE 2
Analysis of variance table for fitted equations

Degrees of Freedom	Source		Sum of squares	
9	Full 2nd degree equation (equation 5)	(equation 9)	371.4	357.4
		Remainder		14.0
9	Residual			29.5
18	Total after eliminating mean		400.9	

The sum of squares due to the 'full second degree equation' (equation 5) has nine degrees of freedom associated with the nine independent constants fitted in addition to the mean. The 'canonical equation' (equation 9) contains only four independent constants apart from the mean. These are two constants of equation (9) and two of the co-efficients in (9a) (the third is fixed by the requirement that the squares of these coefficients sum to unity). We see that the simpler expression is associated with a sum of squares of 357.4 and that the fitting of an equation containing five more constants accounts only for a further 14.0 of the sum of squares.

The estimates in equations (9) and (9a) are not linear functions of the observations and consequently the number of constants in this equation and the number of extra constants associated with the remainder sum of squares cannot be directly associated with degrees of freedom in the usual sense. However the analysis serves to show that the simpler expression probably accounts for the data as well as does the more complicated one.

If we put Y equal to its maximum value in (10) we get $X_1 = 0$ which, when substituted in (11) or (11a), gives a plane of alternative conditions

on which Y attains its maximum value. In general on substituting a lower value for Y in equation (10) we obtain two values for X_1 equal in magnitude but opposite in sign which when entered in (11) give the equations of parallel planes of lower yield 'sandwiching' the maximum plane as illustrated in figure 3. Sections of this system for three levels of the concentration variable are shown on the right hand side of this figure.

At the time when these experiments were carried out (some five years ago) the number of chemical yield surfaces which had been approximately determined was small and this surface, showing such marked dependence between all the variables, was somewhat unexpected. Further experiments having confirmed the reality of the system it was realised that this was a case where the reaction was sufficiently simple to allow a theoretical study which, as it turned out, explained the type of surface found.

Although most chemical systems are more complicated, the study serves to show that, as a result of the laws which govern chemical systems, ridge surfaces of one sort or another are to be generally expected (as experience has in fact confirmed).

3. THE THEORETICAL STUDY

The chemical system could be represented by the following sequence of competitive reactions

$$2a + bNb \rightarrow aNb + ab \qquad \text{(reaction 1)}$$

$$2a + aNb \rightarrow aNa + ab \qquad \text{(reaction 2)}$$

The substance bNb contained a large molecular nucleus N to which were attached the two groupings b. In the part of the sequence denoted by 'reaction 1', one of the b groupings was replaced by an a to form aNb which was the required product. However, as is shown in 'reaction 2', under the conditions in which the first reaction could take place aNb could destroy itself by combining with more a to form the unwanted product aNa.

The concentration of the starting material bNb was kept constant throughout the experiments, the concentration which was varied being that of the substance a. The substance ab was chemically inert and no reverse reaction took place.

If the reactions were allowed to continue for a time t a mixture of unchanged a and bNb and of the products aNb, aNa and ab was produced. The required product aNb had to be separated from the other products in subsequent purification stages.

To proceed we need to use some simple chemical ideas. As they may be unfamiliar to those readers who are not chemists we shall briefly explain them as they are needed.

Concentrations of reactants.

The chemical equations above indicate the proportions in which the actual molecules combine. It is convenient to measure the amounts of the various substances in 'gram moles', one gram mole being the molecular weight in grams of the substance concerned. Thus the equation for reaction (1) implies that two gram moles of a combine with one gram mole of bNb to form one gram mole of aNb and one gram mole of ab. We shall be concerned with the *concentrations* of the reactants in the solvent and these will be expressed as 'gram moles per litre'. The following symbols are used to denote the concentrations and fractional yields of the various reactants in the system.

	Reactant	Conc. at time t	Conc. Initially	"Fractional" yield at time t = (concentration)/c_{20}
Starting materials	a	c_1	c_{10}	η_1
	bNb	c_2	c_{20}	η_2
Products	aNb	c_3	0	η_3
	ab	c_4	0	η_4
	aNa	c_5	0	η_5

In addition the symbol C denotes the ratio c_{10}/c_{20} of the initial concentrations of the starting materials. In the discussion of section 2 and in [1] and [3] the concentration of a was measured as a 'percentage' which was denoted by c. Since the concentration c_{20} was kept constant in our experiment c and C are alternative measures of the concentration of a. In fact $C = 0.4555c$.

We can now derive a theoretical expression for the yield η_3 of aNb in terms of T the temperature, C the relative concentration of a, and t the time, which can be directly compared with the empirical equation.

Since the number of gram moles of a present in the system *in some form or other* at any time t must be constant and equal to c_{10} it follows that

$$c_1 + c_3 + c_4 + 2c_5 = c_{10} \qquad (12)$$

Applying the same reasoning to the nucleus N we have

$$c_2 + c_3 + c_5 = c_{20} \tag{13}$$

whence

$$\eta_2 + \eta_3 + \eta_5 = 1 \tag{14}$$

Finally

$$c_3 + 2c_5 = c_4 \tag{15}$$

By subtracting four times equation (13) from equation (12) and adding equation (15) we obtain

$$c_1 = c_{10} - 4c_{20} + 2c_3 + 4c_2$$

or

$$\eta_1 = C - 4 + 2\eta_3 + 4\eta_2 \tag{16}$$

Kinetic Theory.

We now use two simple concepts from the kinetic theory of chemical reactions. The law of 'mass action' states that, in dilute solution, the speed of a chemical reaction is proportional to the molecular concentrations of the substances reacting. Thus in particular if, at time t, p and q are the molecular concentrations of two substances P and Q which are taking part in a non-reversible reaction to form a substance R, such that one molecule of P reacts with one molecule of Q to form one molecule of R, then the rate of reaction (that is the rate of decrease of the molecular concentrations of P or Q or the rate of increase of the molecular concentration of R) is given by

$$-\frac{dp}{dt} = -\frac{dq}{dt} = \frac{dr}{dt} = kpq \tag{17}$$

where r is the molecular concentration of R at time t. Experimental results have shown that the law often holds *approximately* in moderately concentrated solutions such as we consider.

The quantity k occuring in equation (17) is called the *rate constant.* Its magnitude depends on the temperature (T). Study of a variety of chemical reactions has shown that the relationship between k and T is represented fairly satisfactorily by the empirical equation due to Arrhenius

$$k = \alpha \exp \{-\beta/(T + 273)\} \tag{18}$$

where α and β are constants depending on the reaction studied.

In the sequence with which we are concerned we denote the rate constant for the first reaction by k' and the rate constant for the second reaction by k and define constants α', β', α and β so that

$$k' = \alpha' \exp\{-\beta'/(T + 273)\} \qquad k = \alpha \exp\{-\beta/(T + 273)\} \qquad (19)$$

Then the ratio $\rho = k'/k$ of the rate constants is given by

$$\rho = \gamma \exp\{-\delta/(T + 273)\} \qquad (20)$$

where $\qquad\qquad \gamma = \alpha'/\alpha \qquad$ and $\qquad \delta = \beta' - \beta \qquad (21)$

The equation of the 'theoretical' surface:

Now reaction (1) occurred as follows:

$$a + bNb \xrightarrow{k'} aNb + b$$

$$a + b \xrightarrow{\infty} ab$$
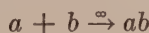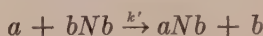
the second part of the reaction being instantaneous.

Similarly for reaction (2)

$$a + aNb \xrightarrow{k} aNa + b$$

$$a + b \xrightarrow{\infty} ab$$
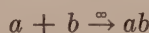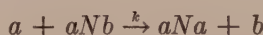
At some particular temperature T then, the rate of disappearance of bNb in reaction (1) is $\rho k c_1 c_2$. Also the rate of formation of aNb from reaction (1) is $\rho k c_1 c_2$ while the rate at which it is destroyed in reaction (2) is $k c_1 c_3$.

We have therefore the pair of differential equations

$$-dc_2/dt = \rho k c_1 c_2 \qquad (22)$$

$$dc_3/dt = \rho k c_1 c_2 - k c_1 c_3 \qquad (23)$$

These together with equations (14), (15), (16) and (19) allow us to obtain expressions for η_1, η_2, η_3, η_4 and η_5, the yields of the products at time t. The derivation is given in section 1 of the appendix. In the particular case of the desired product aNb the yield at time t is

$$\eta_3 = \frac{\rho}{\rho - 1} z_t(1 - z_t^{\rho-1}) \qquad (24)$$

where z_t is a function of T, C, t and c_{20} depending on the constants α, β, γ and δ and defined by

$$c_{20} t \alpha \exp\{-\beta/(T + 273)\}$$

$$= (\rho - 1) \int_{z_i}^{1} z^{-1} \{2(\rho - 2)z^\rho + 2\rho z + (\rho - 1)(C - 4)\}^{-1} \, dz \quad (25)$$

where
$$\rho = \gamma \exp \{-\delta/(T + 273)\} \quad (25a)$$

To the extent that the various assumptions are justified therefore we have an equation for the theoretical yield surface for η_3 and (see equations 74, 76, 77 and 78 of the appendix) incidentally for η_1, η_2, η_4 and η_5 also. In the form in which it is expressed by equations (24) and (25), the characteristics of this surface cannot be readily appreciated so we shall proceed by actually fitting this form of expression to our data and comparing the resulting surface with that obtained empirically. To do this we need to estimate the values of the unknown constants from the data.

The ratio ρ of the rate constants.

Considering again the equations describing the reactions we see that the ratio ρ of the rate constants is the ratio of the probability that an a will replace a b from bNb to the probability that an a will replace a b from aNb.

Now if the chance of an a replacing a b at a particular position on the nucleus N is independent of the type of grouping (a or b) which occupies the other position, then the chance of replacement will be twice as great with bNb, where there are two positions at which replacement can occur, as with aNb, where there is only one. Thus ρ will equal 2 independently of the temperature T.

Now it is readily shown (see section 2 of the appendix) that the maximum yield for aNb is given by

$$\eta_3(\text{max}) = \rho^{-1/(\rho-1)} \quad (26)$$

at which value

$$\eta_2 = \rho^{-\rho/(\rho-1)} \quad \text{and} \quad \eta_5 = 1 - \rho^{-\rho/(\rho-1)} - \rho^{-1/(\rho-1)} \quad (27)$$

Consequently if $\rho = 2$ then η_3 (max) $= 0.5$, and at this maximum value of η_3, $\eta_2 = 0.25$ and $\eta_5 = 0.25$.

The maximum yield actually found was not 50% but about 60%. We must conclude therefore that the probability of replacement of a particular grouping on a half-substituted molecule is not the same as the probability of replacement on an unsubstituted molecule (a conclusion we should in any case expect from chemical considerations). We might however still expect the ratio of these probabilities to be largely independent of temperature, at least over the range we have

considered. This implies that the temperature constants β' and β in equations (19) would be equal and that in equation (20) $\rho = \gamma$ and $\delta = 0$. If we substitute the value $\eta_3(\text{max}) = 0.60$ in (26) we obtain $\rho = 3.4$ and the theoretical distribution of yield between the products bNb, aNb and aNa when η_3 was at its maximum value would then be
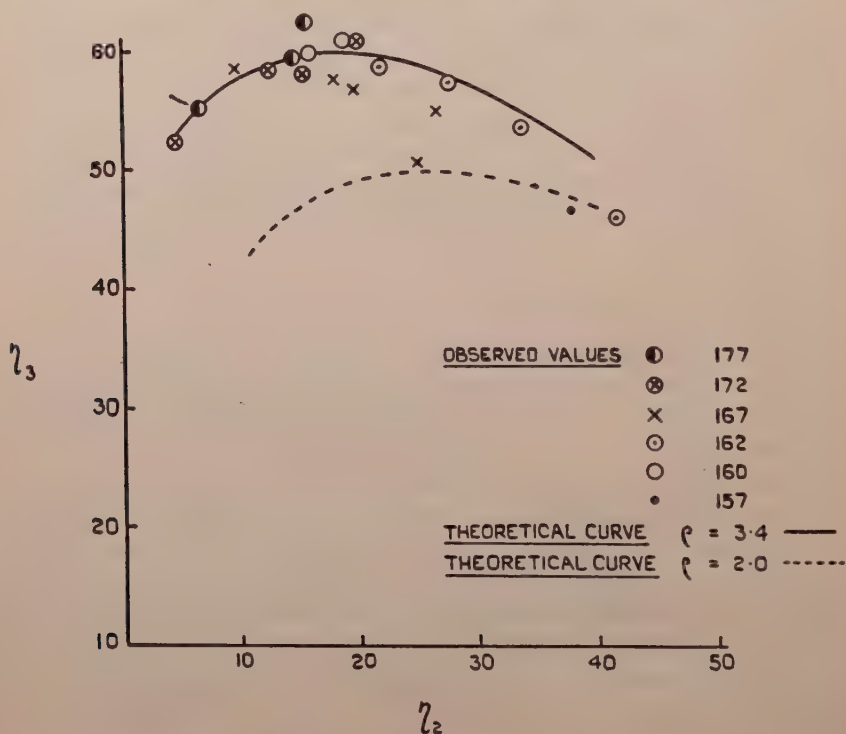
FIGURE 1.   YIELD OF aNb (η_3) PLOTTED AGAINST YIELD OF bNb (η_2) WITH THEORETICAL CURVES FOR $\rho = 3.4$ AND $\rho = 2$.

$\eta_2 = 0.176$, $\eta_3 = 0.600$, $\eta_5 = 0.224$. These values agree quite well with those found in the final four experiments 'on the maximum plane' of the empirical surface for which the averages were

$$\eta_2 = 0.162 \qquad \eta_3 = 0.611 \qquad \eta_5 = 0.213$$

That the value $\rho = 3.4$, is reasonably consistent with the data in other respects can be seen from figure 1 where the observed value $\hat{\eta}_3$ is plotted against $\hat{\eta}_2$. The theoretical relationship between η_3 and η_2 is

$$\eta_3 = \frac{\rho}{\rho - 1} \{\eta_2^{1/\rho} - \eta_2\} \qquad (28)$$

Most of the experimental observations are in fairly close agreement

with the theoretical curve for $\rho = 3.4$ although there seems to be some departure for high values of η_2. Also there seems to be no evidence that the points corresponding to different temperatures follow different lines whose general form is changing steadily with temperature.

The characteristics of our solution are not very sensitive to the value of ρ chosen and we proceed by supposing that ρ is equal to 3.4 and
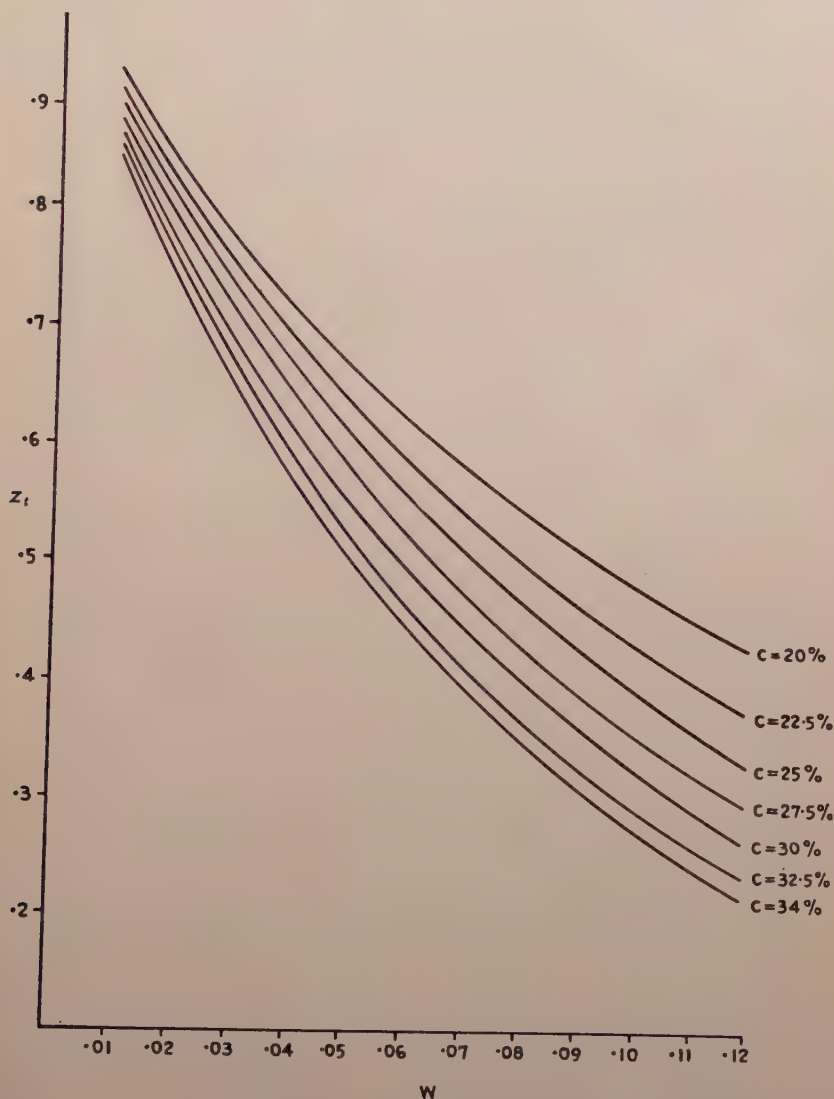
FIGURE 2. GRAPHS SHOWING VALUES OF THE INTEGRAL w FOR VARIOUS VALUES OF z_t AND c.

is independent of temperature. This implies that, the first substitution having occurred, the probability of the second substitution is reduced by a factor of 1.7.

Putting the value $\rho = 3.4$ in (25) we have for our theoretical equation

$$c_{20}t\alpha \exp \{-\beta/(T + 273)\}$$
$$= 2.4 \int_{z_t}^{1} z^{-1}\{2.8z^{3.4} + 6.8z + 2.4(C - 4)\}^{-1} \, dz \qquad (29)$$

Estimates of α and β

The value of the expression on the right-hand side of the equation (29) which we shall denote by $w(z_t, C)$ or more simply by w depends on z_t and C alone or (remembering that the percentage concentration c which we have considered in our experiments is equal to $C/0.4555$) on z_t and c alone. The integral cannot be expressed in terms of elementary functions but its value, for any level of z_t and for the values of c we have employed in our experiments, can be read off from the graphs in figure 2. Each of these graphs was obtained by setting c equal to the appropriate value, calculating ordinates of the curve

$$y = z^{-1}\{2.8z^{3.4} + 6.8z + 2.4(C - 4)\}^{-1} \qquad (30)$$

at the 7 equally spaced values $z = 1.000, 0.875, 0.750, 0.625, 0.500, 0.375, 0.250$ and then calculating the area between $z = 1$ and $z = z_t$ under the curve by numerical quadrature.

The values of the constants α and β can now be estimated from the data as follows. Taking natural logarithms equation (29) may be written in the form

$$u = \ln \alpha - \beta/(T + 273) \qquad (31)$$

where
$$u = \ln w - \ln c_{20} - \ln t \qquad (32)$$

For each experiment the value $\hat{z}_t$ (shown in column 15 of table 1) can be estimated from the formula*

$$\hat{z}_t = \hat{\eta}_2 + \hat{\eta}_3(\rho - 1)/\rho \qquad (33)$$

---

*We notice that in every case the total of $\hat{\eta}_2$, $\hat{\eta}_3$ and $\hat{\eta}_5$ in table 1 is less than the theoretical value of 100%. This is due partly to difficulties of accurate determination of aNa in the presence of other substituents and partly due to some degradation of this product. Because of uncertainty concerning the estimate $\hat{\eta}_5$, z_t was calculated from $\hat{\eta}_2$ and $\hat{\eta}_3$ alone. In references [1] and [3] an empirical surface for aNa was fitted and a region was shown in the maximal plane of aNb where less than 20% of aNa was obtained. In the theoretical equations the yields of both products depend only on z_t and consequently for any surface for which the yield η_3 of aNb was constant the yield η_5 of by-product aNa would be constant also. The region found in the empirical study where aNa was less than 20% probably occurred because degradation of this product was favoured by reaction conditions in this neighbourhood. The effect of this degradation is not allowed for in the theoretical study.

or putting $\rho = 3.4$

$$\hat{z}_t = \hat{\eta}_2 + 0.706\hat{\eta}_3 \tag{34}$$

The corresponding values of $\hat{w}$ for each experiment may now be obtained from the values of z_t by reading from the appropriate graph in figure 2. Finally the values of $\hat{u}$ shown in column 16 of table 1 may be obtained by substituting values of t and $\hat{w}$ for each experiment in (32) remembering that the value for c_{20} was kept constant at 3.10.

From the form of equation (31) we see that we may now obtain estimates of $\ln \alpha$ and β by fitting a regression line of $\hat{u}$ on $x = (T + 273)^{-1}$ by the method of least squares.

We find

$$\ln \hat{\alpha} = 16.86 \pm 3.62 \tag{35}$$

$$\hat{\beta} = 10,091 \pm 1,595 \tag{36}$$

The quantities following the plus and minus signs in (35) and (36) are the formal 'standard errors' calculated in the usual way from the residual sum of squares. It is clear from inspection of the table that the deviations from the regression line contributing to this sum of squares still contain components due to c and t indicating that the theoretical expression does not give a perfect fit to the data. This is not surprising first because of the assumptions we have had to make in the derivation and second because the levels assumed for time t and concentration c are not entirely appropriate. Doubt concerning the level of t exists because, in addition to the reaction occuring during the 'time on temperature', some reaction will also occur while the reaction vessel is being heated up and this is difficult to allow for. The value of c may not be entirely appropriate because owing to solubility factors the effective concentrations in the solvent may be slightly different at different temperatures and at different stages of the reaction. In spite of these limitations a reasonably close representation of the experimental data is achieved by the theoretical expression (29) which contains only three adjustable parameters (ρ, α and β) as compared with the ten adjustable parameters (β_0, β_1, $\cdots$, β_{23}) of the empirical expression.

Comparison of the theoretical and empirical surfaces.

Using our estimates, $\hat{\alpha}$ and $\hat{\beta}$, we may now calculate the value of z_t and hence the values of η_2, η_3 and η_5 predicted by our 'theoretical' equation for any desired level of T, c and t. Figure 4 shows the contours of the resulting 'theoretical surface' for comparison with Fig. 3.
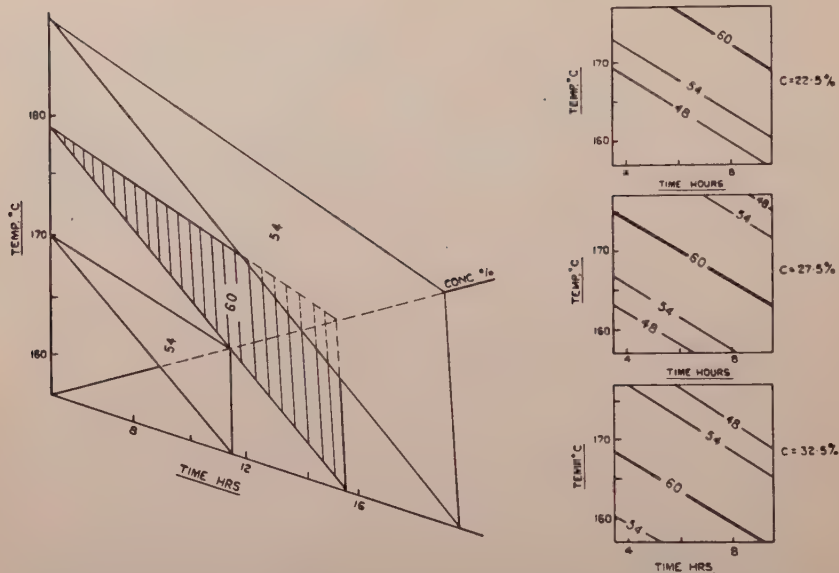
FIGURE 3.   CONTOURS OF EMPIRICAL YIELD SURFACE WITH SECTIONS AT THREE
LEVELS OF CONCENTRATION
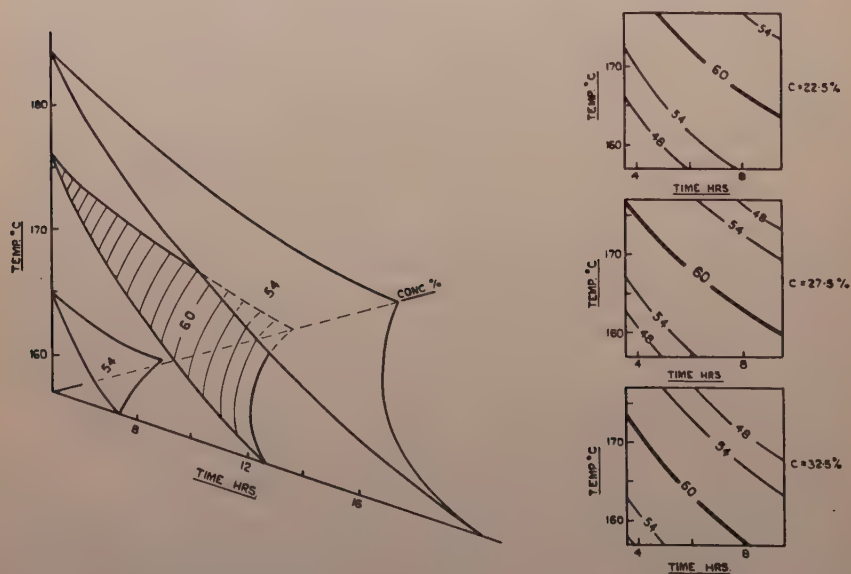
FIGURE 4.   CONTOURS OF THEORETICAL YIELD SURFACE WITH SECTIONS AT THREE
LEVELS OF CONCENTRATION. ρ INDEPENDENT OF TEMPERATURE

It is seen that there is remarkably close agreement between the general characteristics of the two surfaces one of which has been obtained entirely empirically and the other derived on a particular theory of the mechanism of the reaction. In each case there is a whole surface rather than a unique point on which the maximum yield is obtained surrounded on either side by surfaces of lower yield.

A final point of some interest is that z_t and hence the yield η_3 is in fact a function of *four* quantities T, C, t and c_{20} which determines the 'overall concentration' of the reactants. Of these only the first three were varied in the experiments, and c_{20} was kept constant. Had we included c_{20} as a variable which would have been a perfectly sensible thing to do, then the yield surface would still have been completely described in terms of a *single* canonical variable and there would have been a redundancy of three variables instead of two.

An Analogy.

A simple picture* of what is occurring can perhaps be gained from the following analogy.

Imagine a billiard table on which a number of black and white balls are in continuous motion. Suppose that when a black ball collides with a white ball a blue ball is produced and that when a black ball collides with a newly formed blue ball a red ball results.

In this analogy black and white balls correspond to the molecules of the starting materials a and bNb, blue balls to molecules of the required product aNb, and red balls to the unwanted aNa. Starting off with a given number of black and white balls we can see that as soon as the system is set in motion blue balls will begin to appear and these in turn give rise to an increasing number of red balls. Provided that initially there is a sufficient excess of black balls the number of blue balls will increase to a maximum then fall off until finally only red balls and excess black balls remain.

Clearly the proportions of the various sorts of balls on the table at a given instant will depend on the following variables:

(1) The time which has elasped since the start.
(2) The speed with which the balls move.
(3) The relative numbers of black to white balls at the start.
(4) The total number of white balls at the start.

---

*This is intended to provide only a very rough parallel which may be of assistance to non-chemist readers. The real mechanism of chemical reactions is known to be much more complicated and cannot be accounted for by simple collision. For example only a small proportion of molecular 'collisions' actually result in reaction and this proportion is dependent on temperature.

These correspond to the variables t, T, C and the overall concentration c_{20} (temperature T being linked to speed by the Ahhrenius equation). Suppose for fixed conditions of (2) (3) and (4) the time is noted for the maximum proportion of blue balls to be produced. If now conditions (3) and (4) are kept the same but the speed with which each ball moves is doubled the effect will be like that of showing a cinematograph film at twice the rate, an identical sequence of events will be gone through twice as fast and in particular the *same* maximum proportion of blue balls to the initial number of white balls will be produced but in half the time. Similarly if conditions (2) and (3) are kept constant but the initial number of black and white balls on the table is doubled and if we ignore the effect of interference then again a similar sequence of events will occur but at twice the speed and again the *same* maximum proportion of blue balls to the initial number of white balls will be produced but in half the time.

The effect of change in factor (3), the relative number of black and white balls is less easily appreciated intuitively. However we can see that since the *relative* rates at which white balls are disappearing and red and blue balls appearing is at any stage completely independent of the number of black balls (since any change in the number of black balls effects both these rates proportionally) it follows that the proportion of blue balls relative to white balls and the proportion of red balls relative to white balls follows precisely the same course whatever the number of black balls. Consequently once again the same maximum proportion of blue balls to white is produced whatever the proportion of black to white balls. It is evident that for such a system a maximum can be obtained for almost any level of a particular variable provided the other three variables are suitably adjusted.

4. THE CANONICAL VARIABLE

The part played by z_t in the theoretical equation is exactly parallel to that played by the canonical variable X_1 in the empirical equation. This is seen most clearly if we consider a case where z_t may be obtained as an explicit function of T, c, and t. That is to say a case where the integral w in equation (25) may be explicitly evaluated. We have noted already that on the simplest view of the reaction we would expect the ratio ρ of the rate constants to equal 2, and it is readily seen from the form of equations (24) and (25) that, apart from the maximum yield of η_3 being 50% rather than 60%, the general characteristics of the resulting surface will be the same with this value as they are with the value $\rho = 3.4$. Taking $\rho = 2$ we have

$$\eta_3 = 2z_t(1 - z_t) \tag{37}$$

where (see section 3 of the appendix) z_t is now explicitly defined in terms of T, c and t by the equation

$$z_t = (C - 4)/\{C \exp [c_{20}(C - 4)t\alpha \exp \{-\beta/(T + 273)\}] - 4\} \qquad (38)$$

Now subtracting η_3 (max) $= 0.50$ from both sides of (37), and writing Y for $100 \ \eta_3$ (to agree with the notation of the empirical surface) we have

$$Y - Y(\text{max}) = 200z_t(1 - z_t) - 50 \qquad (39)$$

$$= -\{7.07(2z_t - 1)\}^2 \qquad (40)$$

Writing $\dot{W} = 7.07 \ (2z_t - 1)$ we see that the theoretical surface is completely described by the pair of equations

$$\left\{ \begin{array}{l} Y - Y(\text{max}) = -\dot{W}^2 \qquad (41) \\[2mm] \dot{W} = 7.07\left\{\dfrac{2(C - 4)}{\{C \exp [c_{20}(C - 4)t\alpha \exp \{-\beta/(T + 273)\}] - 4\}} - 1\right\} \qquad (42) \end{array} \right.$$

Now (equation 10) the empirical surface is closely approximated by

$$Y - Y(\text{max}) = -3.80X_1'^2 \qquad (43)$$

where substituting $C = 0.4555c$ in (11a) we have

$$X_1' = 0.150T + 0.419C + 0.303t - 31.969 \qquad (44)$$

If we write $W = (3.80)^{\frac{1}{2}}X_1'$ we see that the empirical surface is approximately described by the equations

$$\left\{ \begin{array}{l} Y - Y(\text{max}) = -W^2 \qquad (45) \\[2mm] W = 0.292T + 0.817C + 0.591t - 62.320 \qquad (46) \end{array} \right.$$

which are directly comparable with (41) and (42).

The 'theoretical canonical variable' $\dot{W}$ is a more complicated function of T, C and t than is the empirical canonical variable W. The latter is necessarily a simple *linear* function of T, C and t, and consequently contour surfaces of constant yield in figure 3 are necessarily planes. However over the regions considered these planes do provide a reasonable approximation to the curved contour surfaces of Figure 4 as (for the reason given in Section 1) we might expect them to.

In the discussion above we have compared the canonical variable of the empirical surface with the 'theoretical canonical variable' arising in the particular case when $\rho = 2$. When $\rho = 3.4$ a similar situation will exist and although η_3 will not be a quadratic function of z_t yet a quadratic function will still closely approximate the true curve near the maximum.

In our derivation we have supposed that the ratio $\rho = k'/k$ of the rate constants, was itself independent of temperature. To put it another way we have supposed that the temperature constants β' and β of equation (19) were equal. This supposition is supported by the data over the range of values studied. It is interesting however to consider how the surface would be affected if this were not true and this is perhaps best done by considering an example. Let us suppose that, at the temperature $157°C$, ρ was equal to 2 and that, at the temperature $177°C$, ρ had increased to 3.4. Substituting these values in equation (20) we find that this implies that $\gamma = 12.63$ and $\delta = 5.132$. If we suppose that the values for the constants α and β were the same as before we then have

$$\ln k' = 29.49 - 15{,}223/(T + 273) \tag{47}$$

$$\ln k = 16.86 - 10{,}091/(T + 273) \tag{48}$$

$$\ln \rho = \ln k' - \ln k = 12.63 - 5{,}132/(T + 273) \tag{49}$$

The solid contour model for the (yield: temperature, concentration, time) surface which would then be found is shown in Figure 5 together with sections taken at various levels of concentration. The diagrams
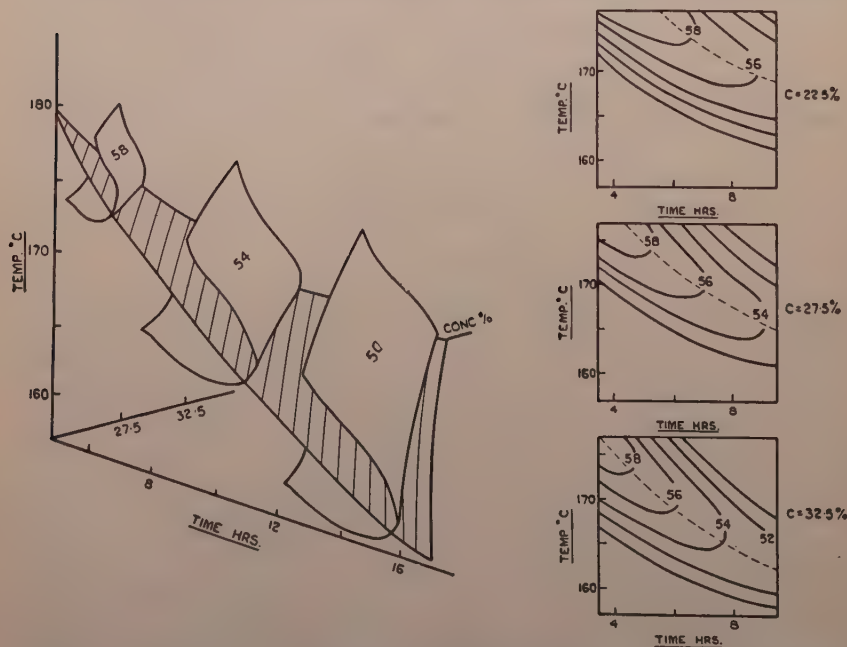
FIGURE 5. CONTOURS OF THEORETICAL YIELD SURFACE WHEN ρ DEPENDS ON TEMPERATURE.

were prepared by carrying through the numerical integration and
subsequent calculations as before for each of the concentrations 22.5%,
27.5%, and 32.5% and for five values of ρ corresponding to five values of
the temperature calculated from equation (49) as follows:

$$T = 157 \qquad 162 \qquad 167 \qquad 172.5 \qquad 177$$

$$\rho = \quad 2.00 \qquad 2.34 \qquad 2.62 \qquad 3.00 \qquad 3.40$$

The three concentration sections were then prepared by drawing smooth
contour lines through the points calculated at these temperatures and
finally from these sections the representation of the solid model was
obtained.

Considering this solid model we see that we now have a situation
where there is a 'rising' ridge instead of a stationary one. The value of
the yield steadily increases on the ridge as the temperature is increased.
A section of the solid model for a particular value of time or concen-
tration gives a two-dimensional rising ridge system running diagonally
to the axes of the variables like the concentration sections shown. A
section of the solid model for a particular value of temperature on the
other hand gives a two dimensional *stationary* ridge system of the type
considered before.

In general we must expect a surface of this sort to occur in the
common case where a competitive system is influenced by a highly
dependent set of variables and one competitor is favoured by a certain
direction of movement in the variables.

In the example we have considered the rising ridge results from
the first reaction in the sequence being favoured by high temperatures.
It is easy to imagine other examples of this sort of phenomenon. For
instance in some systems the rates of competing reactions depend on
different powers of the concentration terms (see for example reference
[5]). In these circumstances a rising ridge associated with concentration
would be expected.

It is of some interest to consider the behaviour of the empirical
method when a rising ridge of this sort occurs. The typical situation
encountered is as follows. Analysis of the fitted second degree equation
yields a canonical equation

$$Y - Y_S = B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2 \tag{50}$$

in which (as with the stationary ridge system discussed in section 2)
one of the coefficients (say B_{11}) is negative and comparatively large
and the other two (B_{22} and B_{33}) are small. The centre S of the fitted
system is remote from the design. To determine the nature of the

system in the region where it applies (in the neighborhood of the design) a new origin S', situated as closely as possible to the design centre and in the plane $X_1 = 0$ is taken. Suppose this new origin is at the point $X_2 = X_{2S'}$ and $X_3 = X_{3S'}$ in the plane $X_1 = 0$. Then writing $X_2' = X_2 - X_{2S'}$ and $X_3' = X_3 - X_{3S'}$ for the new coordinates and substituting these in equation (50) we have

$$Y - Y_{S'} = B_{11}X_1^2 + B_2X_2' + B_3X_3'(+ B_{22}X_2'^2 + B_{33}X_3'^2) \qquad (51)$$

Where B_2 and B_3 measure the slopes of the yield surface *on the plane of the ridge* and will not be negligible if the ridge is non-stationary. If we can ignore B_{22} and B_{33} as negligible in comparison with B_{11}, equation (51) is that of a system having contour surfaces which are parabolic cylinders like that in Figure 11(f) of [1] or Figure 11.8(E) of [3].

We see from equation (51) that, if we wish to move in a direction so that $Y - Y_{S'}$ is made as large as possible, we should, (since B_{11} is negative) make the contribution of the first term equal to zero by keeping $X_1 = 0$ (by remaining on the plane of the ridge). Also we should proceed so that the contribution of the terms B_2X_2' and B_3X_3' is as large as possible. For movement through a given distance r on the plane this will be achieved by following the direction of steepest ascent. Thus X_2' and X_3' should be varied in proportion to B_2 and B_3. We see that this movement which would be at right angles to the yield contours on the plane of the ridge would in the present example lead in the *direction of rising temperature.*

In general where a competitive system is affected by a single factor like temperature or concentration this type of analysis will be helpful in identifying the factor responsible. At the same time it should be borne in mind that in the presence of a ridge system unequivocal identification by this means is not possible. For instance in the present example we see from figure 5 that we could attribute the effect found to the joint influence of time and concentration instead of to temperature. As always it is necessary to consider evidence of this kind in the light of possible theoretical explanations for the phenomenon observed.

6. CHOICE OF METRICS FOR VARIABLES

In the experiments described the standardised variables x_1, x_2, x_3 of the design were linearly related to the natural variables T, c and t as follows $x_1 = (T - 167)/5$, $x_2 = (c - 27.5)/2.5$, $x_3 = (t - 6.5)/1.5$. The unit change of a given variable in the design, taken as a percentage of the average departure of the variable from its natural origin we may call 'the coefficient of unit change', U. Remembering that the natural origin for temperature is $-273°C$, in the present case we have

$$U \text{ (Temp.)} = 100 \times 5/440 = 1.1\%$$

$$U \text{ (conc.)} = 100 \times 2.5/27.5 = 9.1\%$$

$$U \text{ (time)} = 100 \times 1.5/6.5 = 23.1\%$$

When experiments are to be conducted with the intention of fitting an empirical response surface doubt may exist as to whether we should relate the standardised variables of the design to the natural variables by a linear scale, as was done in this experiment, or by a log scale, or a reciprocal scale, or in some other manner.

When the coefficients of unit change are small the surface plotted in terms of transformed variables, like those above, will usually be almost the same in appearance as when plotted in terms of the untransformed variables, since the relationships over the ranges studied between transformed and untransformed variables will be almost linear in this circumstance. Even so by appropriate choice of metrics the interpretation of the fitted equation may be greatly simplified as will be illustrated in section 7.

Theoretical Surface in terms of the New Metrics.

In the present example we have seen (equation 25) the important part which is played by the function

$$c_{20} t \; \alpha \; \{\exp - \beta/(T + 273)\} \tag{52}$$

in describing the yield surfaces. In fact the time t, the *overall* concentration c_{20}, and (if ρ is independent of temperature) the temperature T enter the theoretical equation only through this expression. Its logarithm is

$$\ln c_{20} + \ln t + \ln \alpha - \beta/(T + 273) \tag{52a}$$

which is a linear expression in functions of T, c_{20} and t. If therefore we use a reciprocal scale for absolute temperature and logarithmic scales for the time and the overall concentration the contours of the ridge system will appear as planes in the space of these variables. In figure 6(a) a section of the theoretical yield surface already given in Figure 4 is shown with time plotted on a log scale and temperature on a reciprocal scale.

When ρ is assumed to be temperature dependent, T enters the expression (25) on the right hand side as well as on the left. However as will be seen from figure 6(b), over the ranges considered, the ridge is again rendered almost straight by plotting on the basis of reciprocal absolute temperature and log time. From these diagrams it will be seen that, when the variables are scaled in terms of these new units, a
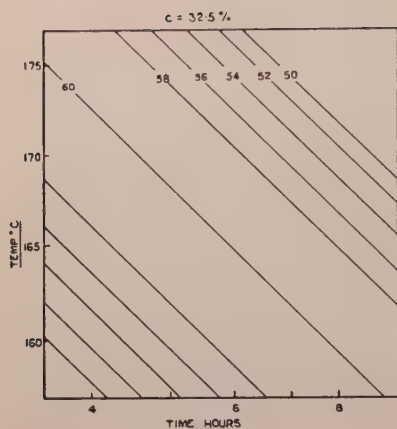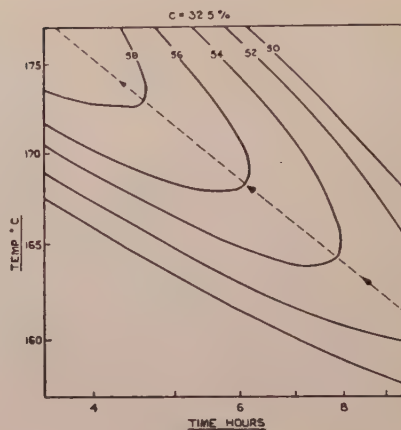
FIGURE 6(a) FIGURE 6(b)

CONTOURS OF THEORETICAL YIELD SURFACES WITH TIME PLOTTED ON A LOG
SCALE AND TEMPERATURE ON A RECIPROCAL SCALE.

considerably closer fit might be expected from a second degree equation.

For many other chemical reactions the compound variable in (52) will play an equally important part in the response function and in the absence of other evidence the choice in empirical investigations of reciprocal scales for absolute temperature and log scales for time and overall concentration would seem to be indicated. One would expect that, on these scales of measurement, the system could be more precisely represented by a simple equation.

The choice, in the present example, of an appropriate metric for C, the concentration of reactant a relative to that of reactant bNb, is less easily decided. However, if we consider the particular case where $\rho = 2$ the equation of the surface for z_t may be written

$$\ln w = \ln \alpha - \beta(T + 273)^{-1} + \ln c_{20} + \ln t$$

$$= \ln \ln \left[\frac{(4z_t + C - 4)}{Cz_t} \right] - \ln (C - 4) \qquad (53)$$

We are particularly interested in the region of the surface where η_3 takes its maximum value. Here $z_t = \frac{1}{2}$. Substituting this value in equation (53) we have for the equation of the surface on which η_3 is a maximum

$$-\beta(T + 273)^{-1} + f(C) + \ln t + \ln \alpha + \ln c_{20} = 0 \qquad (54)$$

where

$$f(C) = \ln \left\{ \frac{C - 4}{\ln (2C - 4) - \ln C} \right\} \tag{55}$$

would seem to be an appropriate metric for C.

Empirical Surface in Terms of the New Metrics.

The expectation that an equation of second degree might fit the data more closely when the variables were expressed in terms of the new metrics $(T + 273)^{-1}$, $f(C)$, and $\ln t$, is borne out, as is seen from the analysis of variance in Table 3. Since, in this example, the co-efficients of unit change are not large no dramatic reduction in the residual sum of squares is to be expected however.

TABLE 3

Analysis of Variance before and after transformation of the variables

Source	DF	Sums of squares	
		Original Metrics	New Metrics
Due to Regression (after elimination of the mean)	9	371.4	380.2
Residual	9	29.5	20.7
Total (after elimination of the mean)	18	400.9	400.9

In refitting the equation after changes in the metrics use may be made of the estimates already obtained in the following way. When recoding the data for the new metrics we need only ensure that the coded data are *linearly related* to the chosen functions. We can there-fore arrange matters so that the coded values of the independent variables $\dot{x}_1$, $\dot{x}_2$, $\dot{x}_3$ are "close" to those of the original independent vari-ables x_1, x_2, x_3. In the present example this was done by arranging that the re-coded levels of the variables for the first eight experiments (the 2^3 factorial part) were -1 and $+1$ as before. For example, for temperature we require a coding $\dot{x}_1 = a + b (T + 273)^{-1}$ such that $\dot{x}_1 = -1$ when $T = 162°C$ and $\dot{x}_1 = 1$ when $T = 172°C$. Substituting these values in the equation and solving we obtain $a = 88, b = 38,715$ whence the coding used for temperature was

$$\dot{x}_1 = 88 - 38,715(T + 273)^{-1} \tag{56}$$

Proceeding in a similar way with the remaining variables we have

$$\dot{x}_2 = -27.9299 + 9.9965 f(C) \tag{57}$$

$$\dot{x}_3 = -7.84868 + 4.25532 \ln t \tag{58}$$

These recoded values are shown in columns (8), (9) and (10) of table 1. The differences from the original coding are not very large and we can therefore regard the coefficients b_0 , b_1 , b_2 etc. already obtained as *first approximations* to the new coefficients $\dot{b}_0$, $\dot{b}_1$, $\dot{b}_2$ etc. Accurate values of $\dot{b}_0$, $\dot{b}_1$, $\dot{b}_2$ etc. were obtained by writing down the normal equations (after elimination of the mean*) for the new recoded variables, inserting b_1 , b_2 , b_3 etc. as first approximations and then obtaining successively closer and closer approximations by "one at a time" and "steepest ascent" relaxation (see for example [7] and [8]). If the elements of the new inverse matrix are required these can be obtained, (for example, by Hotelling's method [9]), using the elements of the known inverse from the original coding as the first approximation.

7. BASIC CONSTANTS AND CANONICAL VARIABLES

We find for the newly fitted equation

$$\dot{Y} = 59.15 + 2.00\dot{x}_1 + 1.01\dot{x}_2 + 0.67\dot{x}_3 - 2.00\dot{x}_1^2 - 0.72\dot{x}_2^2$$
$$- 1.00\dot{x}_3^2 - 2.78\dot{x}_1\dot{x}_2 - 2.18\dot{x}_1\dot{x}_3 - 1.16\dot{x}_2\dot{x}_3 \tag{59}$$

On comparison with equation (5) it will be seen that, as would be expected, the coefficients are quite close to those obtained before.

The canonical form of the equation is also similar. We have

$$\dot{Y} - 59.51 = -3.50\dot{X}_1^2 - 0.41\dot{X}_2^2 + 0.19\dot{X}_3^2 \tag{60}$$

where
$$\dot{X}_1 = 0.760\dot{x}_1 + 0.473\dot{x}_2 + 0.446\dot{x}_3 - 0.329 \tag{61}$$

On decoding we now have an expression for the canonical variable $\dot{X}_1$ in more appropriate functions of the natural variables.

$$\dot{X}_1 = -29{,}423(T + 273)^{-1} + 4.728 f(C) + 1.897 \ln t + 49.839 \tag{62}$$

Putting $\dot{X}_1 = 0$ we see that on the new scales the empirically fitted equation gives for the plane of maxima

$$-29{,}423(T + 273)^{-1} + 4.728 f(C) + 1.897 \ln t + 49.839 = 0 \tag{63}$$

Now we have seen (54) that for $\rho = 2$ the theoretical equation of the

---

*By fitting the equation in the form of $y - \bar{y} = b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + \cdots + b_{11}(x_1^2 - \overline{x_1^2}) + \cdots$, etc. the convergence of the iteration is speeded up.

maximum plane is

$$-\beta(T + 273)^{-1} + f(C) + \ln t + \ln \alpha + \ln c_{20} = 0 \qquad (64)$$

with which (63) may be compared.

If the theoretical system exactly fitted, the coefficients of $f(C)$ and $\ln t$ would be equal and on dividing equation (63) by this common coefficient (63) and (64) would be exactly comparable. Here to enable comparison to be made, we divide (63) through by the geometric mean 2.995 of the coefficients 4.728 and 1.897 of $f(C)$ and $\ln t$ to obtain

$$-9,824(T + 273)^{-1} + 1.579f(C) + 0.633 \ln t + 16.641 = 0 \qquad (65)$$

Comparison of (65) and (64) shows that the canonical variable X_1 is *carrying as coefficients the constants of the reaction.* The value 9,824 is an estimate of β and (since $\ln c_{20} = 3.10$) $16.64 - 3.10 = 13.54$ is an estimate of α. (Both are in reasonable agreement with the estimates of equations 35 and 36).

The lack of equality of the coefficients of $f(C)$ and $\ln t$ does not support theoretical expectation. However, this is probably because $\rho \neq 2$ and also because for reasons given already, the 'effective reaction times' are greater than the values assumed and the 'effective relative concentrations' are less. It would be possible to calculate appropriate 'correction factors' which could indicate how our basic theory should be modified. We shall not pursue this topic here however.

This example served to point out to us two interesting possibilities which have been borne in mind and developed in later investigations. These are:

1) Where sufficient is known of the nature of the basic mechanism (i.e. the kinetics in chemical examples) we may proceed to fit a surface based on this mechanism rather than on the empirical Taylor series.
2) When we start off with little knowledge of a system careful study of the characteristics of a fitted empirical surface, particularly as elucidated by canonical analysis, can lead to a *conception* of the probable basic mechanism. A first guess can then be tested and improved upon by a process of 'experimental iteration'.

Of these (2) is possibly the more important and may have applications outside chemistry, for example, the characteristics of the surface for a fertilizer trial in agriculture or a nutrition experiment in biology might supply important information on the metabolism of the plant or animal cell.

8. BASIC CONSTANTS AND THE DIRECTION OF STEEPEST ASCENT

We have seen above that, in the example we have studied, essential information concerning the reaction constants was contained in the coefficient of the canonical variable X_1.

It is perhaps worth noting also that the direction of steepest ascent would also contain much of this information. For we see that if the true surface could be represented in terms of a single canonical variable so that

$$\eta - \eta(\max) = \lambda(px_1 + qx_2 + rx_3 + s)^2 \tag{66}$$

then multiplying out this expression and equating the coefficients to the constants $\beta_0, \beta_1, \cdots, \beta_{23}$ we have

$$\left.\begin{array}{llll}
\beta_0 = \eta \max + \lambda s^2 & \beta_1 = 2\lambda sp & \beta_2 = 2\lambda sq & \beta_3 = 2\lambda sr \\[4pt]
\beta_{11} = \lambda p^2 & \beta_{22} = \lambda q^2 & \beta_{33} = \lambda r^2 \\[4pt]
\beta_{12} = 2\lambda pq & \beta_{13} = 2\lambda pr & \beta_{23} = 2\lambda qr
\end{array}\right\} \tag{67}$$

Thus for this type of example the constants $\beta_1, \beta_2,$ and β_3 which define the direction of steepest ascent are in fact proportional to the coefficients p, q, r in the canonical variable, as is at once obvious from the consideration that the direction of steepest ascent is at right angles to the contour plane of maxima in the space of the factors.

In examples like the present one estimates of $\beta_1, \beta_2, \beta_3,$ and β_{12}, β_{13} and β_{23} are available after the first eight experiments. If these are such as would support the hypothesis that

$$\frac{\beta_{12}}{\beta_1\beta_2} = \frac{\beta_{13}}{\beta_1\beta_3} = \frac{\beta_{23}}{\beta_2\beta_3} \tag{68}$$

we may begin to suspect (although we are entitled to do no more) that we may be dealing with a system having a single dominant canonical variable.

9. SOME REMARKS ON THE PROCESS OF SCIENTIFIC INVESTIGATION

The technique of scientific investigation contains two essential processes

a) the devising of experiments suggested by the investigator's appreciation of the situation to date and designed to elucidate it further;

b) the examination of results of experiments performed to date in the light of all background knowledge available, with the object of postulating theories susceptible of test in future experimentation.

The first is essentially a movement from 'theory' to 'experiment' indicated in Figure 7 by an arrow pointing upwards, the second is a

movement from 'experiment' to 'theory' indicated by an arrow pointing downwards.

During a complete investigation these processes of synthesis and analysis used in alternation will normally be employed many times and, by what we may call 'experimental iteration', the investigator should be led closer and closer to the truth.

Most investigations first pass through a 'speculative' stage. Here statistical methods can rarely be of help but it is nevertheless vital that this early work should be done fully and with imagination, otherwise later effort may be wasted in detailed investigation of the wrong

FIGURE 7. DIAGRAMATIC REPRESENTATION OF PROCESS OF EXPERIMENTAL ITERATION.

basic system. Statistical methods provide efficient tools for investigating a system whose general nature has been broadly decided. They provide no substitute for basic scientific thinking about what the system to be investigated should be. It is the duty of a statistician to dissuade the experimenter from employing these methods until he has done sufficient preliminary work to decide what basic system he should explore more fully and incidentally until he has acquired reasonable skill in carrying out experiments with the system.

To appreciate the interplay of processes (a) and (b) let us imagine the beginning of a chemical investigation. At stage 1 in Figure 7, the experimenter would have some, perhaps not very precise, idea as to the general way in which some chemical might be manufactured. Process (a) would begin in his mind something like this—"I believe that in suitable circumstances reactant A would combine with reactant B to form C. From theoretical knowledge, my own experience, and other people's experience of similar reactions cited in the literature, I should think that conditions X might be worth trying".

The appropriate experiment would then be performed (stage 2 in

the diagram). As soon as the results were seen the second type of mental process, denoted by (b) above would start—"The reaction did produce a little of the desired product C but there was a very large amount of unwanted product D also present. This could be due to the large amount of water which had to be used to dissolve the reactants and which would favour formation of D".

He has now reached stage (3) at which point the first kind of mental process (a) begins again—"If I carry out the reaction using a non-aqueous solvent I may avoid the large production of by-product D". He is thus led to perform a further experiment at stage (4) using a non-aqueous solvent, and so on.

When the speculative experiments have led to some reasonably well defined system which is sufficiently promising to justify development much will be gained by using the powerful tools provided by applied mathematics such as "steepest ascent", empirical surface fitting and "theoretical surface study".

It should be noted that these techniques still employ the basic processes (a) and (b), and that our applied mathematics helps as much with (b) as it does with (a). Thus we are not only concerned with designing experiments which will estimate the 'effects of the factors' (process a) but also with making calculations (for example of the direction of steepest ascent, and of the canonical form of a fitted equation) which suggest what further experimentation should be performed (process b).

There has been a tendency for some statisticians to concentrate on the, perhaps rather rare, experimental situation where a single group of experiments is planned and from the result some irrevocable decision is to be made. Such an investigation is concerned exclusively with a single application of procedure (a). It is customary to emphasise in such a situation the danger of taking action on a hypothesis suggested by inspection of the data, but not in mind when the experiment was planned.

This point has sometimes been misunderstood and interpreted to mean that process (b) was in some way suspect and should not be indulged in. In fact of course it is fundamental to investigation which would be quite barren without it. In researches of the type we have discussed the experimenter and the statistician should examine the data most carefully and any hypothesis which appeared possible and important should be submitted to the test of later experimentation.

The particular sequence of techniques shown in Figure (7) is not to be regarded as providing a set pattern which should always be followed. The use of these and other devices will be decided by such circumstances

as the degree of basic knowledge concerning the mechanism of the system and the object and importance of the study. If, for example, the experimenter were required merely to make in the laboratory a few pounds of some rare organic chemical for some special research purpose, he would be quite content to do a few speculative experiments sufficient to allow him to prepare this small amount of material in reasonable quality, the finding of an economic process would not be worth the trouble. At the other end of the scale, for a large and expensive manufacture an elaborate and costly study would be justified.

10. ACKNOWLEDGEMENT

APPENDIX A—DERIVATION OF CERTAIN RESULTS QUOTED IN THE PAPER

1) *General solution of the differential equations.*
   From equations (22) and (23) we have

$$dc_3/dc_2 = \rho^{-1}c_3/c_2 - 1 \tag{69}$$

Write $c_3 = c_2 s$, then

$$dc_3/dc_2 = s + c_2(ds/dc_2) \tag{70}$$

Substituting (70) in (69) we have

$$\frac{-dc_2}{c_2} = \frac{ds}{1 + s(\rho - 1)/\rho} \tag{71}$$

$$-\ln c_2 + \text{constant} = \frac{\rho}{\rho - 1} \ln \{1 + s(\rho - 1)/\rho\} \tag{72}$$

Now when $t = 0$, $c_2 = c_{20}$, and $s = 0$ hence the constant in $\ln c_{20}$.
Now $\eta_2 = c_2/c_{20}$ and $s = c_3/c_2 = \eta_3/\eta_2$
Thus

$$\eta_2 = \left(\frac{\rho - 1}{\rho} \frac{\eta_3}{\eta_2} + 1\right)^{-\rho/(\rho-1)} \tag{73}$$

Write

$$\eta_2 = z^\rho \tag{74}$$

After substituting this expression in (73) and rearranging we have

$$\eta_3 = \frac{\rho}{\rho - 1} z(1 - z^{\rho-1}) \tag{75}$$

and using equation (14)

$$\eta_5 = 1 - \frac{\rho}{\rho - 1} z + \frac{1}{\rho - 1} z^\rho \tag{76}$$

From equation (15)

$$\eta_4 = 2 - \frac{\rho}{\rho - 1} z - \frac{\rho - 2}{\rho - 1} z^\rho \tag{77}$$

Finally from equation (16)

$$\eta_1 = C - 4 + \frac{2\rho}{\rho - 1} z + 2 \frac{(\rho - 2)}{\rho - 1} z^\rho \tag{78}$$

Now from equation (22)

$$\frac{-d\eta_2}{dt} = c_{20} \rho k \eta_1 \eta_2 \tag{79}$$

Substituting (78) in (79) and noting that $d\eta_2/dt = \rho z^{\rho-1} dz/dt$ we have after rearrangement

$$-kc_{20} \frac{dt}{dz} = \frac{\rho - 1}{z\{2(\rho - 2)z^\rho + 2\rho z + (\rho - 1)(C - 4)\}} \tag{80}$$

whence

$$kc_{20}t = (\rho - 1) \int_{z_t}^{1} z^{-1} \{2(\rho - 2)z^\rho + 2\rho z + (\rho - 1)(C - 4)\}^{-1} dz \tag{81}$$

and using the Arrhenius equation (18) we obtain equation (25). Equations (74), (75), (76), (77) and (78) together with (81) yield the complete solution allowing the yields η_1, η_2, η_3, η_4 and η_5 to be evaluated at any time t.

2) *Maximum yield of aNb.*

   If we put $dc_3/dt = 0$ in equation (23) we have

$$\eta_3/\eta_2 = \rho \tag{82}$$

substituting in (73) $\eta_2 = z^\rho = \rho^{-\rho/(\rho-1)} \tag{83}$

Whence

$$\eta_3 = \rho^{-1/(\rho-1)} \tag{84}$$

which is readily shown to be a maximum.

At this point · $\eta_5 = 1 - \rho^{-1/(\rho-1)} - \rho^{-\rho/(\rho-1)} \tag{85}$

3) *Solution when $\rho = 2$.*

In the particular case when $\rho = 2$ equation (81) becomes

$$kc_{20}t = \int_{z_t}^{1} z^{-1}(4z + C - 4)^{-1} \, dz \tag{86}$$

giving

$$kc_{20}t = \{\ln (4z_t + C - 4) - \ln (Cz_t)\}/(C - 4) \tag{87}$$

After rearrangement this yields the explicit function for z_t

$$z_t = (C - 4)/[C \exp \{c_{20}(C - 4)tk\} - 4] \tag{88}$$

whence using the Arrhenius equation (18) we obtain equation (38).

REFERENCES

[1] Box, G. E. P. The Exploration and Exploitation of Response Surfaces: Some general considerations and examples. *Biometrics, 10:* 16, 1954.

[2] Box, G. E. P. and Wilson, K. B. On the Experimental Attainment of Optimum Conditions. *J. R. Stat. Soc., B, 13:* 1, 1951.

[3] *The Design and Analysis of Industrial Experiments*, Box, G. E. P., Connor, L. R., Cousins, W. R., Davies, O. L. (Editor), Himsworth, F. R., and Sillito, G. P., London: Oliver and Boyd, 1954.

[4] Arnold, D. S., Box, G. E. P., Erickson, E. E., Hunter, J. S., Nelli, J. R., Pike, F. P. The application of Statistical Procedures to the Flooding Capacity of a Pulse column, Progress Report No. 3 Contract No. AT-(40-1)-1320.

[5] Erickson, E. E., Hunter, J. S., Nelli, J. R., Pike, F. P. The flooding capacity of a Pulse Column on the Benzene-Water System. Progress Report No. 4. Contract No. AT-(40-1)-1320.

[6] Hammett, L. P. *Physical Organic Chemistry, 1st Ed.*, 1940. New York: McGraw-Hill.

[7] Fox, L. A short account of Relaxation Methods. *Quart J. Mechanics and Applied Mathematics, 1:* 253, 1948.

[8] Booth, A. D. An application of the Method of Steepest Descents to the Solution of systems of Non-Linear Simultaneous equations. *Quart. J. Mechanics and Applied Mathematics, 2:* 460, 1949.

[9] Hotelling, H. Some new methods of Matrix Calculation, *Ann. Math. Stat., 14:* 1, 1943.

DESIGN AND ANALYSIS OF TWO PHASE EXPERIMENTS

G. A. McIntyre

Division of Mathematical Statistics,
C.S.I.R.O., Canberra, Australia

Introduction

It sometimes happens in experimental work that the effects of different treatments cannot be measured directly and a further stage of testing is required in order to evaluate them. Examples of this type of situation are studies of the effect of conditions of growth of parent material on resistance to disease or productivity of progeny; the survival of nodule bacteria under various conditions of storage and appraised by inoculating appropriate legume seedlings; and the effect of various treatments on virus multiplication in leaf tissue, the concentration of virus being ascertained by lesion counts on indicator plants.

Principle of Design

In order to have a measure of consistency of performance, due to treatment, and a valid basis for a test of significance it is essential that there should be replication in the first phase. Further, it is essential that the product of each plot of the first phase should be separately evaluated in the second phase.

Replication in the second phase is not necessary but is highly desirable where uncontrollable variation in this phase is large relative to anticipated effects. The comparison of sugar cane varieties in sugar content per unit weight involves a replicated variety trial in the first phase followed by chemical analysis of the product of each plot, this perhaps being done in duplicate or triplicate. Replication in the second phase is in this case usually done only as a check against analytical mistakes. On the other hand if one wished to determine the effect of various spacings between plants in a seed crop on the production per acre in the following generation then with anticipated small or no effects of spacing one would probably (a) replicate the testing of seed from each plot of the first phase and (b) employ a design in the second phase which would enable the elimination of major changes of soil fertility and still permit a valid analysis of the total yields for seed deriving from each first phase plot.

It is the purpose of this note to examine the relation of second to first phase designs to achieve this objective. Only those designs will be considered for which the total or mean yields of plots in the second phase which derive from a plot in the first phase can be analysed directly in accordance with the first phase design. This implies that errors of measurement for material from each treatment plot should be uniform for which it is necessary that replication in the second phase shall be constant.

Designs

Arrangements satisfying the above requirements can be classified in various ways depending on the particular aspect considered important. However it is doubtful if classification is of much assistance and instead a series of examples will be given to illustrate the principle of design. It is not claimed that even in their general form these examples exhaust the possible arrangements.

(1) First Phase: Any replicated design

Second Phase: The plots of the first phase are regarded as varieties in the second phase and are incorporated in any design for which the error variance for comparison of every variety pair is uniform.

Thus material from plots of a 5×5 latin square in the first phase could be assayed in a balanced lattice square design for 25 varieties in the second phase. The weighted means with recovery of interblock information would be entered into the latin square for analysis of effects, the latin square error from this analysis being appropriate to the comparison of treatments.

A special and common instance of this class of design occurs when the second phase replicates of material from plots of the first phase are completely randomised. There need not be more than one plot in the second phase corresponding to each treatment plot of the first phase. As a particular example one could cite the second phase testing of survival of nodule bacteria using clover seedlings grown in test tubes under conditions so nearly standardised that there is little or nothing to be gained over complete randomisation. The chemical analysis associated with the comparison of sugar cane varieties could also be classed as a further instance. If there is an unavoidable time lag in analysis with a consequent effect on the assay then the simplest valid procedure would be to test the products of the various first phase plots completely at random or in random order within a stratum of the first phase. The latter method of course would be potentially more efficient.

(2) First Phase: Any stratified design

Second Phase: The plots from each stratum of the first phase

design are assayed separately from other strata but the same form of design and degree of replication is used for all strata. The design should be such that the error variance for every comparison of pairs within all strata is the same.

For example suppose that six treatments are compared in the first phase in a randomised block arrangement with four replicates. The material from each replication could be assayed in the second phase using a 6×6 latin square. Here squares are confounded with first phase blocks while rows and columns within squares are orthogonal to treatments within blocks so that effects associated with strata in neither the first nor second phases contribute to treatment and error mean square in the analysis.

(3) First Phase: Any stratified design

Second Phase: One or more repetitions of the design with one-to-one correspondence of plots in the first phase and assay plots in the second phase. Subsequently there would be randomisation of strata and plots within strata in the second phase. Thus if the numbered plots of a latin square are re-randomised by rows and columns to give a design for the second phase, then material from a particular numbered plot in the first phase would be tested on the plot of corresponding number in the second phase. In this class of design there is complete confounding of stratification between the two phases.

(4) First Phase: 6×6 latin square

Second Phase: Randomised block with 6 plots and 6 treatments. The material from the six plots in a column of the latin square is assigned at random to plots in a block of the second phase. This is a degenerate case of the preceding example.

It is of interest to note that if the designs of the first and second phases were reversed the row stratification of the second phase is not confounded with stratification of the first phase nor orthogonal to treatments within blocks. Analysed as a randomised block the error variance but not the treatment mean square would be inflated by the row effects of the second phase and in fact for a valid analysis the data would have to be analysed according to the second phase design.

(5) This example illustrates the possibilities of more elaborate arrangements, not desirable in themselves, except where there is a potentially worthwhile economy in the use of test material to attain a given level of precision.

The following unpublished data by courtesy of D. J. Goodchild, Plant Pathology Laboratories, University of Sydney, relates to an investigation of the effect of four light treatments on the synthesis of tobacco mosaic virus in leaves of tobacco *Nicotiana tabacum* var.

Hickory Pryor. Four successive leaves at defined positions on the stem were taken from each of eight plants of comparable age and vigor after inoculation with buffered and diluted sap expressed from infected tobacco plants. The eight plants were topped and kept under a constant and continuous light intensity for 48 hours prior to inoculation.

Arbitrarily grouping the plants into two sets of four, the four treatments were applied to the leaves, which had been separated from the plants and were sustained by flotation on distilled water, in a latin square design to each set with plant source as columns and leaf positions as rows.

After treatment, virus content of each leaf was assayed by expressing sap, diluting with phosphate buffer to an appropriate dilution and inoculating half leaves of the assay plants, *Datura stramonium*, on which countable lesions appeared. Dilutions from leaves belonging to the first column in the latin squares of each set were regarded in effect as eight treatments which were assayed in a 4 × 4 graeco-latin design using half leaves at four consecutive positions on four assay plants, treatments from a column within a set belonging to the same alphabet. Similarly for the leaves belonging to the second, third and fourth columns of the first phase sets.

In Table 1(a) the plan of assignment of treatments to plants and leaves within plants is given together with a plot number which is used to identify the source of virus for the half leaves of Table 1(b). Included also in 1(b) are the square roots of counts which were transformed

<div align="center">

TABLE 1(a)

First Phase: Two 4 × 4 latin squares

</div>

Leaf Position	TEST PLANTS				Leaf Position	TEST PLANTS			
	1	2	3	4		5	6	7	8
1	a 90.8 1	b 116.7 5	c 84.9 9	d 64.4 13	1	a 61.5 17	b 69.1 21	c 76.2 25	d 64.5 29
2	b 66.9 2	a 49.7 6	d 77.3 10	c 72.7 14	2	c 89.8 18	d 88.1 22	a 54.4 26	b 55.3 30
3	c 91.2 3	d 92.5 7	a 75.5 11	b 56.5 15	3	d 101.5 19	c 84.7 23	b 81.7 27	a 64.0 31
4	d 85.4 4	c 91.5 8	b 83.2 12	a 60.7 16	4	b 78.6 20	a 78.0 24	d 71.6 28	c 68.1 32

<div align="center">

Treatment Totals   a 534.6   b 608.0   c 659.1   d 645.3

</div>

TABLE 1(b)

Second Phase: Four 4 × 4 graeco-latin squares

ASSAY PLANTS

Leaf Position	1		2		3		4	
1	1	24.9	2	12.7	3	17.5	4	16.1
	17	18.5	20	12.3	18	17.2	19	19.3
2	2	24.5	1	23.1	4	22.5	3	25.8
	18	31.6	19	24.3	17	11.9	20	18.2
3	3	30.5	4	22.2	1	27.7	2	14.7
	19	32.3	18	22.8	20	23.7	17	13.7
4	4	24.6	3	17.4	2	15.0	1	15.1
	20	24.4	17	17.4	19	25.6	18	18.2

ASSAY PLANTS

Leaf Position	5		6		7		8	
1	5	28.0	6	11.5	7	26.5	8	17.1
	23	18.0	22	21.3	24	17.8	21	11.1
2	8	21.7	7	25.8	6	10.4	5	25.7
	22	23.6	23	22.2	21	15.1	24	18.7
3	7	20.1	8	32.0	5	34.9	6	13.2
	21	18.6	24	26.7	22	20.2	23	22.3
4	6	14.6	5	28.1	8	20.7	7	20.1
	24	14.8	21	24.3	23	22.2	22	23.0

ASSAY PLANTS

Leaf Position	9		10		11		12	
1	9	23.2	10	15.4	11	12.7	12	17.3
	28	20.3	25	14.8	27	14.6	26	14.9
2	10	24.2	9	14.7	12	17.2	11	25.8
	27	25.7	26	11.7	28	15.6	25	17.4
3	11	20.8	12	21.5	9	19.5	10	22.0
	26	16.1	27	23.2	25	17.1	28	18.7
4	12	27.2	11	16.2	10	15.7	9	27.5
	25	26.9	28	17.0	26	11.7	27	18.2

ASSAY PLANTS

Leaf Position	13		14		15		16	
1	13	13.8	14	18.4	15	13.1	16	9.4
	30	9.2	31	21.9	29	12.1	32	6.8
2	16	15.2	15	16.7	14	21.5	13	13.3
	31	17.0	30	13.6	32	19.2	29	16.9
3	15	15.4	16	21.1	13	18.1	14	17.2
	32	22.3	29	18.0	31	13.8	30	14.6
4	14	15.6	13	19.2	16	15.0	15	11.3
	29	17.5	32	19.8	30	17.9	31	11.3

to equalise variance within treatments, and the sums of counts of the same identification number are recorded in 1(a). Thus the total for the first leaf of Plant 5 in the first phase is 61.5, given by the sum of 18.5, 11.9, 13.7 and 17.4.

So far as the leaves of either latin square set are concerned the plots within columns are tested in latin squares in the second phase. This is a design of the type given in Example (2) and the contribution to latin square error and treatment and row mean squares within sets is the within alphabet variance of the graeco-latin square. However the latin column mean square is inflated in addition by the between graeco-latin squares contrast for a single alphabet.

It is not obvious however that the contribution to the mean square

for treatments taken over both alphabets is also the within alphabet variance. The contrast between treatments a, b is the contrast of plots identified as 1,17 and 2,20 in the first graeco-latin square and so on. If Δ be a measure of deviation from the norm for a particular whole leaf and ϵ for a half leaf within a whole leaf then the elements of contrast from these sources of variation in the first graeco-latin square are

(1)	(17)		(2)	(20)
$\Delta_{11} + \epsilon_{11a}$	$\Delta_{11} + \epsilon_{11b}$		$\Delta_{12} + \epsilon_{12a}$	$\Delta_{12} + \epsilon_{12b}$
$\Delta_{22} + \epsilon_{22a}$	$\Delta_{23} + \epsilon_{23b}$	minus	$\Delta_{21} + \epsilon_{21a}$	$\Delta_{24} + \epsilon_{24b}$
$\Delta_{33} + \epsilon_{33a}$	$\Delta_{34} + \epsilon_{34b}$		$\Delta_{34} + \epsilon_{34a}$	$\Delta_{33} + \epsilon_{33b}$
$\Delta_{44} + \epsilon_{44a}$	$\Delta_{42} + \epsilon_{42b}$		$\Delta_{43} + \epsilon_{43a}$	$\Delta_{41} + \epsilon_{41b}$

where the subscripts give the row and column position respectively in this first square. Similarly for the remaining graeco-latin squares. Collecting coefficients of the different elements of variation, squaring, adding and dividing by 64, which is the number of contrasted half leaves for these treatments in the second phase, the expectation of the contribution to variance from these sources is $\sigma_\Delta^2 + \sigma_\epsilon^2$. The within alphabet variance which can be derived from the graeco-latin square analyses is an estimator of this.

As the contribution of variance from the graeco-latin squares to treatments is the same within and over the two sets, it follows that the interaction of treatments with sets also has this component of variance and therefore the pooling of this interaction with the latin square error within sets in the conventional analysis of duplicated latin squares is an unbiased procedure. The only source of variance in the analysis of the sets which does not include a contribution from the graeco-latin squares equal to the within alphabet variance is the mean square for the set contrast.

The formal analysis of the two primary sets is given in Table 2. Included also is a pooled analysis of the graeco-latin squares, following Yates, to provide estimates of second phase components.

Components of variation contributing to the mean squares are listed. They are, in order

σ_V^2 — variance between treatment means
σ_{TP}^2 — variance between test plants
σ_{TR}^2 — variance between leaf positions in test plants
σ_L^2 — residual variance between leaves of test plants
σ_{AP}^2 — variance between assay plants

σ_{AR}^2 — variance between leaf positions in assay plants

σ_{A}^2 — residual variance between whole leaves of assay plants

σ_{ϵ}^2 — variance between halves within leaves of assay plants

The above notation is strictly appropriate for random variates. With constants for treatments and for positions within test and assay plants and with only additive effects, σ_V^2, for example, stands for $\sum_1^4 (V_a - \overline{V})^2/(4 - 1)$.

The treatment mean square is significant at the 5% level. The means of treatments in half leaf units are a, 16.71; b, 19.00; c, 20.60; d, 20.17 with S. E. of $(30.25/32)^{\frac{1}{2}}$ or 0.97.

For the assay plants the mean squares for plants and leaf position are significant at the 1% and 5% levels respectively. For the test plants the mean square for plants contains a component due to differences between assay plants not present in the error mean square. The effect of leaf position on test plants is not statistically significant.

By equating the mean squares to their expectations expressed in components, estimates of the variances of these components may be determined. Thus the estimate of σ_A^2 is $\frac{1}{2}$ (11.80–7.60). For σ_{TP}^2 the sum of squares for sets and plants within sets were pooled and likewise the coefficients of variances of components corresponding to these sums of squares. The estimate of σ_{TP}^2 was then determined by elimination of the contributions from other components using the estimates of these from other mean squares. For the majority of estimates the error of estimation will be considerable. The estimated values given at the foot of Table 2 will however be used in the discussion for illustrative purposes.

Discussion

Modification of Design

Provided that there is replication at the second phase so that an estimate of the variances associated with the various stratifications can be estimated in both phases as in the numerical example, it is possible to predict from the data of one or more trials the likely consequences of modifying the design. The several arrangements which follow have been chosen to illustrate the type examples given earlier and are correspondingly numbered. In each instance a treatment is assayed by 32 half leaves in the second phase so that the variances of treatment means are the corresponding error variances in half leaf units divided by 32.

1. First Phase: Two 4 × 4 latin squares

    Second Phase: Complete randomisation of four assays from each first phase plot within the 128 half leaf positions.

TABLE 2

	Degrees of Freedom	Mean Square	F	Expected Value of Mean Square							
				σ_V^2	σ_{TP}^2	σ_{TR}^2	σ_L^2	σ_{AP}^2	σ_{AR}^2	σ_Δ^2	σ_e^2
Latin Squares											
Sets	1	41.40			16		4			1	1
Plants within sets	6	96.44			16		4			1	1
Leaf Position within sets	6	63.45	2.10				4			1	1
Treatments	3	97.23	3.21*			16	4	4		1	1
Error	15	30.25		32			4				1
Graeco-Latin Squares											
Sums Plants within squares	12	75.12	6.37**			2		8		2	1
Leaf Position within squares	12	35.77	3.03*			2			8	2	1
Latin letters within squares	12	44.34		2			2			2	1
Greek letters within squares	12	32.48		2			2			2	1
Error	12	11.80								2	1
Differences											
Alphabets	4	18.07			16		4				1
Latin letters within squares	12	24.12		2		2	2				1
Greek letters within squares	12	12.91		2		2	2				1
Error	36	7.60									1
Estimated variances of location components					1.97	2.07	5.14	7.91	3.00	2.10	7.60

*Significant at the 5% level
**Significant at the 1% level

2. First Phase: Randomised blocks, 4 treatments within each of 8 plants

Second Phase: Treatments within plants of the first phase are assayed in random order in each of two plants in the second phase, using both halves of a leaf for the same virus source.

3. First Phase: Two 4×4 latin squares

Second Phase: Each first phase latin square is duplicated in the second phase with one-to-one correspondence, using both halves of a leaf for the same virus source.

4. First Phase: Two 4×4 latin squares

Second Phase: Treatments within plants of the first phase are assayed in random order in each of two plants in the second phase, using both halves of a leaf for the same virus source.

The error variances in half leaf units for these modifications are expressed in terms of the components and estimates are given by substituting the estimates of variance from the numerical analysis.

Modifi-cation	Components of Error Variance							Estimated error variance
	σ^2_{TP}	σ^2_{TR}	σ^2_L	σ^2_{AP}	σ^2_{AR}	σ^2_A	σ^2_ϵ	
1			4	120/127	96/127	126/127	1	39.98
2		4	4		2	2	1	46.64
3			4			2	1	32.36
4			4		2	2	1	38.36

All of these modifications are less efficient than the actual experimental design.

Modification of amount of replication

This is a particular case of modification of design. Factors limiting increase in replication in the two phases and relative cost in time and labour may be quite different. In the experimental situation discussed in Example 5 it is not physically convenient to increase replication in the first phase beyond eight plants for any one trial but the effect of changes in the amount of replication in the second phase could be considered. The error variance for treatments is not changed if the design is modified by using adjacent pairs of latin square columns to be the alphabets of a graeco-latin square in the second phase. This makes possible a comparison between variances of treatment means for four, six and eight graeco-latin squares in the second phase.

Second phase	Treatment replication in half leaves	Components of Variance							Estimated error variance	Estimated variance of treatment means
		σ^2_{TP}	σ^2_{TR}	σ^2_L	σ^2_{AP}	σ^2_{AR}	σ^2_Δ	σ^2_ϵ		
4 squares	32			4			1	1	30.26	0.946
6 squares	48			6			1	1	40.54	0.845
8 squares	64			8			1	1	50.82	0.794

The gain in information by doubling the work in the assay phase is only 19%. With duplication of the trial as designed by repetition in time there would be a gain of 100%.

Failure to replicate treatments in the first phase

There is a psychological hazard in two phase experiments that the second phase will not be recognised as only an assay. The position is strictly analogous to the difficulty sometimes experienced in realising that sampling variation within plots is not of direct relevance to the comparison of treatments for which the appropriate error variance is the replicate error.

The point scarcely requires illustration but one could consider the consequences of the following procedure with material of the type used in the numerical example.

First Phase: One replication of four treatments using four leaves from one plant.

Second Phase: First phase treatments assayed in random order in each of 16 plants of the second phase, using both halves of a leaf for the same virus source.

Components of Variance					Estimated Error Variance	
σ^2_{TR}	σ^2_L	σ^2_{AR}	σ^2_Δ	σ^2_ϵ	ignoring first phase components	including first phase components
32	32	2	2	1	17.80	248.52

The error variance for the comparison of the assayed leaves is 17.80. The true error variance for the comparison of treatments is 248.52, but this would not be ascertainable from the results of the experimental procedure as no estimate would be available of the average effect of leaf position on the test plant and the residual leaf effect. To

identify differences between the four leaves with treatment effects only and to use the error variance of 17.80 as the basis of comparison would be wrong in principle and could lead to quite misleading conclusions in practice.

Summary

In experiments where the effects of treatments have to be determined by a subsequent stage of assaying it is important to consider design in the assay phase in relation to the design in the primary phase. Examples of designs which enable the assay totals deriving from material in first phase plots to be analysed for treatment and error mean square according to the first phase design are discussed.

REFERENCE

Yates, F. The design and analysis of factorial experiments. Imperial Bureau of Soil Science Technical Communication No. 35.

ERRATUM

KIMBALL, A. W.   Short-cut formulas for the exact partition of χ^2 in contingency tables.   Vol. 10, No. 4 (1954), 452–458.

At the bottom of page 454, in the formula for χ_k^2, the N in the numerator should appear outside of the curly brackets instead of inside.

THE BRADLEY-TERRY PROBABILITY MODEL AND PREFERENCE TASTING

N. T. Gridgeman

Division of Applied Biology
National Research Council,
Ottawa, Canada

INTRODUCTION

Broadly speaking, two methods of appraising sensory differences may be distinguished: scoring and ranking. There are also two sources of sensory difference: that among the intensities of several stimuli identical in kind, and that among preferences in a group of stimuli that may or may not be of the same kind. This paper is concerned with the applicability of the simplest form of ranking, namely, pair comparisons, to testing in general and taste preferences in particular.

In organoleptic work it is usually rewarding to postulate a sensory continuum whose points, S, are monotonically related to the concentration, C, of a given stimulus in a given medium. Some controversy has centred on the meaningfulness of this notion, and on its right to come within the ambit of metrology at all; here, without discussion, the view will be adopted that the concept is operationally valid, and that the practical problem is to refine the measuring technique. A common further assumption is that the relation approximates to the Fechnerian form $S = \alpha + \beta \log (C/C_0)$, where α and β are constants and C_0 is the threshold detectable concentration, over a certain critical range. We shall not at the moment perpend this relation, but may observe that, whatever the true equation, its parameters are likely to be biological variants over the universe of tasters.

A preference continuum is a more nebulous concept. In simplest form, it may be thought of as a series of points forming an ordinate of preference, P, to an abscissa of concentration, C, of a given stimulus; and it is not difficult to visualize a curve that maximizes P at some particular concentration. But a preference continuum must also be applicable to a series of stimuli that are at least partially different in kind. This necessitates a more complex model in which P is a function of an n-dimensional vector quantity. Can such a continuum be validated? In other words, can a subject's preference statements, in suitably selected and controlled situations, concerning a set of flavors, be

---

Issued as N. R. C. No. 3689.

taken to stem from graded reactions of delectability? Is relative delect-ability quantifiable? These questions are important insofar as prefer-ence itself is technologically important; it is in fact concerned with taste in the everyday usage of the word.

To arrange a given set of flavors in some sort of order whose statistical significance can be assessed, we can restrict attention to pair compari-sons, the pairs forming incomplete blocks of two. This method has the attraction of breaking the test down into the simplest possible decision units, and of not requiring graded responses from the taster. To make best use of the data we need a model of the relation between the probability of individual pair judgements and points on the postu-lated continuum. Such a model has been put forward by Bradley and Terry (1, 2). Its applicability to tasting for relative sweetness of graded concentrations of sugar has already been demonstrated in this laboratory by Hopkins (3). Part of his experimental work was on preferences among primary-flavor mixtures, and these too seemed to fit the model. This work has now been extended to cover a wider flavor range. By way of introduction, a sketch of the Bradley-Terry model (see, also, Thurstone, (4)) may help explain the design and interpretation of the experiment.

THE BRADLEY-TERRY MODEL

The probability of a taster's judging that sample i contains more of the stimulus than sample j will depend on the sensation difference $S_i - S_j$ measured on the continuum. (Merely for expository conveni-ence, we shall argue in terms of stimulus intensity.) If this quantity is large and negative the probability will be small; if zero, the probability will be 0.5; and if large and positive the probability will be close to unity. For a particular pair of stimulus concentrations C_i and C_j , taster-to-taster differences in the probability of the specified judgement will be a reflection of "non-parallelism" among the underlying curves of $S = f(C)$, i.e., the tasters' discriminatory powers in that sensory region must be heterogeneous.

As a first approach, let us assume that the probability density, over all values of $S_i - S_j$ for a particular taster, is normal, so that

$$\Pr \{i > j\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{S_i - S_j} \exp\left(-y^2/2\right) dy$$

However, at the expense of introducing a trivial departure from nor-mality, a simple algebraic model of the mechanism can be set up. The point of departure is the observation that a certain "squared hyperbolic secant" function, which is slightly leptokurtic to the normal function

(and whose integration yields the logistic curve*), has these character-
istics (see Fig. 1):

$$\frac{A}{A + B} = A = \frac{1}{4} \int_{-\infty}^{y} \mathrm{sech}^2 \frac{y}{2} \, dy = \frac{e^y}{1 + e^y}$$
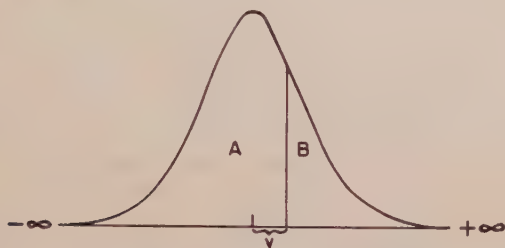
FIG. 1.   THE SQUARED-HYPERBOLIC-SECANT FUNCTION.   THE MAXIMUM ORDI-
NATE IS ¼; THE AREA ENCLOSED IS UNITY.

Hence $y = \ln (A/B)$.   If we now put

$$A = \text{anti-ln } S_i = \pi_i$$

and $$B = \text{anti-ln } S_j = \pi_j$$

we immediately arrive at the position that

$$\Pr \{i > j\} = \pi_i/(\pi_i + \pi_j) \tag{1}$$

and we conceive of π as a function of S—although an *ad hoc* function
in that the two π's sum to unity.   We may at once generalize to the case
of t samples (stimulus concentrations) for which $\pi_1 + \pi_2 + \pi_3 + \cdots$
$+ \pi_t = 1$, and for any pair of which equation (1) applies.   There are
$t(t - 1)/2$ pairs and, therefore, that many probability estimates (as
frequencies of specified judgements in replicate pair comparisons) from
which the π's can be estimated.

These π's, termed "ratings" (although they are of no practical
importance as metrical ratings), have two uses: firstly, they provide a
means of checking the model; secondly, significance tests can be derived
from them.   Maximum likelihood estimates of π are calculated from
the $t(t - 1)/2$ frequency estimates f_{ij}/n of the probabilities $\Pr \{i > j\}$,
the equations (t in number) being

$$f_i/\hat{\pi}_i - n \sum_{j \neq i} (\hat{\pi}_i + \hat{\pi}_j)^{-1} = 0$$

where f_i is the frequency of selection (in a specified direction e.g.,

*Thus the distribution difference is exactly that underlying the logit–probit issue (or, to keep
abreast of current Berksonian terminology, the logit–normit issue).

"the stronger flavored") of i, summed over all pair comparisons. As, by definition, $\sum \pi_i$ is unity, the equations can be solved (by iteration) for the unknowns.

A given set of t frequency sums may arise from any one of many different combinations of the $t(t-1)/2$ individual frequencies, combinations that will differ in fit to the requirements of the rating model. From the model we can estimate the expected ratings, and hence the expectations for the individual frequencies. Finally, from the observed and expected cell frequencies (half each of f_{ij} and f_{ji}), Bradley has shown how to evaluate an index of discrepancy that closely approximates χ^2 with $t(t-1)(t-2)/2$ degrees of freedom.

The use of $\hat{\pi}$ to test for significant sensory differences among the samples stems from the following considerations: The simplest alternative hypotheses are: equality (H_0) *versus* general but undetailed inequality (H_1). If H_0 is true, the probability of each of the possible sets of t frequency sums, as a chance occurrence, can be derived from first principles, although as t and n increase the computational work soon becomes burdensome. All the sets can be listed in such order that their cumulative probabilities provide significance levels for the acceptability of the null hypothesis, an order determined by the likelihood-ratio test statistic

$$B_1 = n \sum_{j \neq i} \log \left(\hat{\pi}_i + \hat{\pi}_j \right) - \sum_t \left(f_i \log \hat{\pi}_i \right)$$

whose minimum is zero and whose maximum is associated with the unit end-point of the probability accumulation. Hence the cumulative probability corresponding to any particular B_1 is that of its not being exceeded if H_0 is true.

Bradley and Terry have tabulated all π, B_1 and the cumulative probabilities for $t = 3$; $n = 1(1)10$, and for $t = 4$; $n = 1(1)6$. These tables are not conspicuously easy to enter as they stand, a defect easily remedied however by the construction of a new (first) column of squariances (sums of squares of deviations from the mean f_i) to serve as a pathfinder. When n is large, say ≥ 20, Bradley and Terry recommend the evaluation of B_1 and the likelihood ratio as usual; then, since the colog of the latter may be regarded as $\chi^2/2$, significances are ascertainable without recourse to special tables. As n increases, the probability that a set of observed frequency sums could arise fortuitously from the null hypothesis ever more closely approximates a smooth monotonic function of the squariances. In other words, χ^2-type curves seem to be approached independently of B_1 and therefore of the ratings and the model. Thus, in the limit, the fit of the model is immaterial to inter-sample significance testing, which can be done directly with

$$\sum_i f_i^2/n - nt(t-1)^2/4$$

regarded as a statistic distributed as χ^2.

On the other hand, the larger the scale of the experiment, the better the test of the model fit will be, so that an investigation into the applicability of the method to small-scale work should be conducted with high n. This condition also enables inter-sample significances to be assessed before the ratings etc. are calculated, thus permitting immediate rejection of undiscriminated sets of judgements, none of which, of course, can yield any information about the fit of the model.

EXPERIMENTAL

Four small flavor modifications were introduced into each of two basic materials, fruit juice and meat. Thus six paired taste contrasts were obtained for each material. Six subjects made 20 replicate preference judgements on each of the 2×6 pairs. Altogether, therefore, 1440 decisions were made.

TABLE I.

Constitution of the Eight Samples
(in parts)

Reference	Fruit juice	Ground meat
A	98 orange + 2 lemon	50 beef + 50 pork + 0.2 salt
B	90 orange + 10 apple	50 beef + 50 pork
		+ 0.1 tenderizer
C	95 orange + 5 lemon	52½ beef + 47½ pork + 0.2 salt
D	85 orange + 5 lemon	52½ beef + 47½ pork
	+ 10 apple	+ 0.1 tenderizer

Canned sweetened orange juice was modified by admixing canned apple juice and lemon juice. The meat samples were formed of blends of ground beef and pork, the additives being salt or a commercial tenderizer; these blendings involved slight textural disuniformities. The composition of the 8 samples is given in Table I. The juices were served in 30 ml. aliquots at 60°F. The meat was formed into one-ounce patties, broiled for 35 minutes, and served hot.

A tasting schedule was drawn up to allow every subject to make a complete set of 12 coded comparisons per day, half each at mid-morning and mid-afternoon sessions. The procedure was repeated for 22 daily sessions, the first two of which, however, were dummy in the sense that,

TABLE II.

Recorded Frequency f_{ij} of Specified Preferences in 20 Replicate Pair Comparisons

Material	Specified sample preference	Subject reference:					
		I	II	III	IV	V	VI
Fruit juices	$A > B$	12	4	9	10	5	14
	$A > C$	11	17	16	8	14	8
	$A > D$	16	13	17	13	13	9
	$B > C$	6	15	17	11	19	3
	$B > D$	10	19	15	13	17	3
	$C > D$	13	7	8	8	11	16
Meats	$A > B$	9	18	12	15	17	18
	$A > C$	9	11	7	10	10	9
	$A > D$	15	17	12	14	18	16
	$B > C$	5	6	4	3	4	9
	$B > D$	7	7	10	10	11	13
	$C > D$	14	14	12	13	13	11

unknown to the subjects, the results were discarded. The order of presentation of any 6 comparisons over the two daily sessions was randomized, with the restriction that, overall, each pair was presented 10 times with one sample on the right-hand side, and 10 times vice versa. Preferences were recorded on simple forms.

RESULTS AND ANALYSIS

All preferences, as frequencies of choice (f_{ij}) in a specified direction, are assembled in Table 2. Summed preferences per sample (f_i) are given in Table 3 together with the corresponding χ^2 $(= \sum f_i^2/20 - 180)$ for the discrimination of each subject for each material.

Homogeneity Tests

Possible inter-subject differences were tested by means of Haldane's treatment of $2 \times N$ contingency tables when some expectations are small (5). The preferences, and their complements to 20, over all 6 subjects, were arrayed for each of the six pairs of both materials, and the 12 values of χ^2 calculated. The first and second moments, k_1 and k_2, of the corresponding conditional distributions of χ^2 for the marginal totals were then obtained from Haldane's formulae, which may be written:

<div align="center">TABLE III.</div>

Summed Preferences Frequencies f_{ij} , and χ^2 (df $= 3$) for Discrimination

Material	Sample	Subject reference:						Total
		I	II	III	IV	V	VI	
Fruit juices	A	39	34	42	31	32	31	209
	B	24	50	43	34	51	12	214
	C	36	15	15	29	18	45	158
	D	21	21	20	26	19	32	139
	$\chi^2 =$	11.7**	36.1**	31.9**	1.7	35.5**	27.7**	34.7**
Meats	A	33	46	31	39	45	43	237
	B	23	15	22	18	18	24	120
	C	40	37	41	40	39	33	230
	D	24	22	26	23	18	20	133
	$\chi^2 =$	9.7*	29.7**	10.1*	18.7**	29.7**	15.7**	96.3**

*: exceeding 5% point of χ^2.
**: exceeding 1% point of χ^2.

$$k_1 = S(N - 1)/(S - 1),$$

and
$$k_2 = \frac{2S^3(N - 1)(S - N)}{(S - 1)(S - 2)(S - 3)} \left(\frac{1}{S - 1} - \frac{1}{AB} \right),$$

where N, the number of subjects, is 6, and S, the total number of judgements, is 6×20, and $A = S - B$ is the sum of all subjects' preferences, in the specified direction, for the particular pair comparison. The moments, summed over each material, were then used to express in standard measure the differences between the observed values of χ^2 and their expectation for homogeneity. The deviates turned out to be 13.1 for the fruit juices and 1.1 for the meats. So it appears that these subjects had remarkably similar tastes for meats, and remarkably dissimilar tastes for the fruit juices.

A plot of the aggregate frequencies over subjects and pairs for each day and each material showed no visible secular trend. To test for general heterogeneity among replicates, Cochran's index of discrepancy

$$Q = (n - 1) \frac{n \sum f_d^2 - (\sum f_{ij})^2}{n \sum f_{ij} - \sum f_{ij}^2}$$

was employed (6). Here n is the number of replicated sets of pair

TABLE IV.

Indices, Q, of Discrepancies Between Replicates in Preferences

Material	Subject reference:					
	I	II	III	IV	V	VI
Fruit juices	20.5	28.9	20.7	16.3	26.3	17.5
Meats	18.1	14.1	18.6	15.0	28.0	29.0

Percentage points of χ^2 for 19 degrees of freedom:
    5%: 30.1; 1%: 36.1

comparisons, i.e. 20; f_d is the number of preferences (in the specified directions) in any one set; and f_{ij} is the number of preferences in the n replicates of any one pair. The limiting distribution of this index is that of χ^2. The Q values are assembled in Table 4, where it will be seen that none of the 12 is high enough to suggest a discrepancy.

The Bradley-Terry Fit

If the kind of preference continuum discussed above is real, we may calculate expected frequencies, as already shown, from the rating estimates. This having been done, goodness of fit of the observed frequencies was tested by Bradley's suggested method (7) of calculating an index of discrepancy $\sum [(f - f')^2 / f']$, the summation extending over all preferences (and their complements to 20) for each subject with each material. For perfect fit, this index is distributed as χ^2 with 3 degrees of freedom. The results are shown in Table 5, from which it is apparent that, in general, the fit is good. Of the 12 values of χ^2 two lie just beyond the 5% point, but as the 6-subject totals, at 19.7 for the

TABLE V.

Values of χ^2 for Goodness of Fit of Preferences to the Bradley-Terry Model

Material	Subject reference:					
	I	II	III	IV	V	VI
Fruit juices	1.81	8.14	1.05	3.74	1.49	3.49
Meats	6.04	2.10	1.60	1.58	3.40	8.11

Percentage points of χ^2 for 3 degrees of freedom:
    5%: 7.82; 1%: 11.35

juices and 22.8 for the meats, are little in excess of the expectation of 18, the two high values are probably fortuitous.

DISCUSSION

Care must be exercised in the interpretation of tests of goodness of fit of the Bradley-Terry model. Even if the model is not the best, it is unlikely to be far from truth, so that, as already intimated, an experiment of normal size could hardly be expected to yield evidence of misfit. Furthermore, the greater the sensory difference induced by the compared samples, the greater the likelihood of very small cell frequencies, with consequent inflation of the index of discrepancy. However, the large-scale trial already made (3) on sweetness intensity has shown excellent agreement with the model, and the present work indicates that the model applies equally well to sensory preferences. The conclusion that a preference continuum, analogous to a continuum of sensation intensity, is, at least in some circumstances, a workable concept, is the most important single outcome of the investigation.

Inter-subject preference variation was of course not unexpected, yet it emerges that there was close agreement about the delectability of the meat samples—those without tenderizer were preferred.

Individual preference for the fruit juices varied, but in general the addition of 5% lemon juice to orange juice was less favored than the addition of 2% lemon juice or 10% apple juice. That there should be more agreement over a desirable quality of meat than over that of fruit juice mixtures is perhaps understandable.

ACKNOWLEDGEMENTS

REFERENCES

1. Bradley, R. A. and Terry, M. E. The rank analysis of incomplete block designs. I. *Biometrika*, 1952, *39*, 324.
2. Terry, M. E., Bradley, R. A. and Davis, L. L. New designs and techniques for organoleptic testing. *Food Technol.*, 1952, *7*, 250.
3. Hopkins, J. W. Incomplete block rank analysis: some taste test results. *Biometrics*, 1954, *10*, 391.
4. Thurstone, L. L. Psycho-physical analysis. *Amer. J. Psychol.*, 1927, *38*, 368.
5. Haldane, J. B. S. The use of χ^2 as a test of homogeneity in a $(n \times 2)$-fold table when expectations are small. *Biometrika*, 1945, *33*, 234.
6. Cochran, W. G. The comparison of percentages in matched samples. *Biometrika*, 1950, *37*, 256.
7. Bradley, R. A. Incomplete block rank analysis: on the appropriateness of the model for a method of paired comparisons. *Biometrics*, 1954, *10*, 375.

SUR LA DETERMINATION DE L'AXE D'UN NUAGE
RECTILIGNE DE POINTS

Georges Teissier

Station Biologique de Roscoff

Au cours d'une étude de la relation d'allométrie développée devant la première Conférence internationale de Biométrie (1948) et parue ici-même, j'ai été conduit à examiner le problème de la liaison linéaire de deux variables jouant des rôles symétriques, telles par conséquent qu'aucune d'elles ne peut légitimement être considérée comme indépendante. Transportant dans le domaine du calcul une technique d'interpolation graphique classique, j'ai proposé de représenter la relation cherchée par la droite qui rend minimum la somme des aires des triangles rectangles ayant cette droite pour hypoténuse commune, deux côtés parallèles aux axes, et leurs sommets aux points X_1, X_2. La forme de l'équation ainsi obtenue:

$$\frac{X_1 - \overline{X_1}}{\sigma_1} = \frac{X_2 - \overline{X_2}}{\sigma_2}$$

incite à écrire, si l'on étudie n grandeurs:

$$\frac{X_1 - \overline{X_1}}{\sigma_1} = \frac{X_2 - \overline{X_2}}{\sigma_2} = \frac{X_3 - \overline{X_3}}{\sigma_3} = \ldots \frac{X_n - \overline{X_n}}{\sigma_n} \tag{D}$$

expression qui donne simultanément, par l'équation d'une droite de l'espace à n dimensions, toutes les relations existant entre les variables prises deux à deux, trois à trois . . . . (1948, p. 54).

Les propriétés de cette droite ont été étudiées depuis par Kermack et Haldane (1950), pour le cas de deux variables, et par Kruskal (1953), pour celui de n variables. Mais, s'il ne semble pas utile de revenir longuement sur le cas de deux variables, il est certain que le problème général mérite d'être repris avec plus de précision. Le présent travail développe une communication sur ce sujet présentée en mon nom par G. Darmois à la troisième Conférence internationale de Biométrie (1953). Un court résumé en a été donné ici-même (1954).

Je rappellerai seulement, pour le cas de deux variables, que le choix d'une ligne de regression traduisant au mieux les résultats d'un ensemble de mesures, dépend des conditions dans lesquelles ces mesures ont été pratiqués, de la nature et de la grandeur de l'incertitude qui pèse sur

ces dernières, et aussi du but du travail poursuivi. Dans bien des cas, l'objet de la recherche est moins de fournir une règle de prévision, que d'extraire des données d'observation une expression approximative de la loi fonctionnelle qui unirait les deux variables comparées, si toutes les causes de perturbations pouvaient être éliminées. C'est ce problème que j'avais tenté de résoudre dans le cas de la relation d'allométrie, sans peut-être avoir montré assez clairement qu'il s'agissait là d'un cas particulier de ce que Quenouille (1952) nomme très justement la recherche de la "loi sous-jacente" à un ensemble de résultats statistiques. Renvoyant à l'excellent livre de cet auteur pour tout ce qui concerne les généralités sur cette question, je rappellerai seulement que, dans le cas de deux variables, les renseignements dont on dispose ne permettent généralement pas de donner une solution unique au problème posé. Il en va autrement, comme nous allons le voir, lorsque l'étude porte simultanément sur un plus grand nombre de variables.

Pour préciser notre recherche, en lui donnant en même temps une forme simple, imaginons que nous ayons pratiqué n mesures sur chacun des éléments d'un échantillon extrait d'une population homogène d'animaux adultes, et supposons que l'espèce étudiée, un Crustacé Brachyoure par exemple, soit l'une de celles pour lesquelles les relations d'allométrie se vérifient bien. C'est dire qu'en moyenne les logarithmes X des mesures varient proportionnellement les uns aux autres et que les corrélations qui les unissent deux à deux sont fortes. Les points figuratifs X_1, X_2 $\cdots$ X_n, constituent dans ce cas un nuage très allongé et sensiblement rectiligne. La détermination de la forme de ce nuage nécessitera, dans l'hypothèse d'une distribution normale, l'estimation des moyennes, des variances et des covariances, soit de $n(n+3)/2$ paramètres distincts, qu'il faudra ensuite combiner diversement pour obtenir des informations utilisables. Mais il est clair que certaines caractéristiques de la distribution dépassent en importance toutes les autres : ce sont celles qui permettent de définir la ligne sur laquelle est axé le nuage. Si celui-ci est très étroit, la connaissance de l'"axe" pourra même suffire à la plupart des applications pratiques. Plusieurs procédés peuvent être employés, pour estimer la position de cette droite, à partir des moments des deux premiers ordres.

Une condition de moindres carrés, imposant à la somme des carrés des distances des divers points figuratifs à la droite cherchés d'être un minimum, conduit à adopter comme solution du problème le plus grand axe de ellipsoïdes d'égale probabilité (Voir par exemple Cramer 1946). Comme cet axe dépend des unités de mesure adoptées, il convient d'user de variables réduites, ce qui ramène notre problème au calcul de la

première composante principale de Hotelling; on détermine cette composante à partir de la matrice des corrélations, par une méthode directe si les variables sont très peu nombreuses, par un procédé d'itération dans le cas général. On sait que la fraction de la variance totale dont rend compte la première composante principale est mesurée par la première racine caractéristique de la matrice et qu'elle est plus grande que celle que l'on pourrait extraire par toute autre fonction linéaire des mesures calculée à partir des mêmes coefficients de corrélation. C'est en ce sens que cette droite (D′) peut être considérée comme étant la droite de meilleur ajustement. Mais on doit noter que si (D′) passe plus près des points observés qu'aucune autre droite, elle ne permet pas, en revanche, de restituer les coefficients de corrélation avec une précision égale à celle que l'on peut obtenir par d'autres procédés.

Le problème peut en effet être abordé d'une autre manière. Si nous admettons *a priori* qu'il existe une droite (D″) représentative du phénomène étudié, nous pouvons convenir d'en donner une représentation paramétrique et écrire, pour chacun des X une relation $X_p - \overline{X}_p = \alpha_p (T - \overline{T})$ où T est une variable auxiliaire, le facteur général, qu'il reste à définir, et α_p le coefficient de regression de X_p en T égal à $r_{Tp} \sigma_p / \sigma_T$, où r_{Tp} est le coefficient de corrélation de X_p et T. Les écarts individuels aux n relations ainsi définies sont indépendants de T; s'ils sont également indépendants entre eux, ce qui n'est nullement nécessaire, mais peut arriver, on doit avoir, pour chacun des $n (n - 1)/2$ coefficients de corrélation une relation de la forme $r_{pq} = r_{Tp} r_{Tq}$ et il s'agit d'estimer, à partir de ce nombre surabondant d'équations, les n coefficients r_{Tp}. Ces saturations en facteur T des n variables sont données par une formule due à Spearman et l'on en déduit immédiatement une définition de T qui permet d'en estimer la valeur pour tout individu. Cette estimation du facteur général est d'ailleurs peu utile, T n'étant qu'un intermédiaire dans des calculs où sa forme analytique n'intervient pas explicitement, et pouvant être éliminé sans inconvénient de l'expression du résultat final qui s'écrira:

$$\frac{X_1 - \overline{X}_1}{r_{T_1}\sigma_1} = \frac{X_2 - \overline{X}_2}{r_{T_2}\sigma_2} = \frac{X_3 - \overline{X}_3}{r_{T_3}\sigma_3} = \cdots = \frac{X_n - \overline{X}_n}{r_{T_n}\sigma_n} \qquad (D'')$$

équation d'une droite (D″) qui diffère à la fois de (D) et de (D′). La solution obtenue épuise toutes les informations dont on dispose sur la variance liée des n grandeurs et restitue, aux erreurs d'échantillonnage près les différents coefficients de corrélation. L'estimation qu'elle donne des différents X est, en revanche, moins précise que celle que permet d'obtenir la droite (D′).

Dans le cas général, les écarts individuels des diverses variables aux lignes de regression de X en T ne sont plus indépendants et un facteur unique ne suffit plus à expliquer les corrélations observées. Le problème est alors d'extraire un facteur général d'un ensemble de données justiciables d'une analyse factorielle. Il peut être résolu par des procédés très semblables à ceux que l'on utilise pour la recherche de la première composante principale. Comme dans le cas d'un facteur unique, la droite (D'') restitue les coefficients de corrélation mieux que la droite (D') et les différents X moins bien qu'elle.

On doit alors se demander quelle est, des droites généralement distinctes (D') et (D'') celle qu'il convient de retenir comme solution au problème posé. Mais la réponse à cette question ne peut pas être immédiate. Les calculs qui donnent les équations de l'une ou de l'autre des deux droites devraient en effet normalement se poursuivre par l'extraction d'autres composantes ou d'autres facteurs qui seraient nécessaires à l'interprétation complète des faits observés. En arrêtant notre analyse à sa première étape, et en limitant notre étude à la recherche d'une droite d'ajustement, nous sacrifions nécessairement une part plus ou moins grande de l'information incluse dans les données dont nous disposons, part qui n'est pas exactement la même dans les deux méthodes que nous comparons. Pour préciser ce point, nous examinerons d'abord quelques cas particuliers.[1]

Pour deux variables, l'analyse factorielle n'est pas utilisable : on ne dispose en effet, pour calculer deux saturations que d'une seule équation de définition $r_{12} = r_{T1}r_{T2}$. Il en résulte que (D'') peut être l'une quelconque des droites comprises entre les lignes de régression de X_2 en X_1 pour laquelle $r_{T1} = 1$ et $r_{T2} = r_{12}$, et de X_1 en X_2 , pour laquelle $r_{T2} = 1$ et $r_{T1} = r_{12}$, la droite (D') superposée à (D) correspondent à $r_{T1} = r_{T2} = \sqrt{r_{12}}$. Ce résultat montre de façon particulièrement claire l'incertitude fondamentale du choix de la relation linéaire unissant deux variables et la nécessité d'une hypothèse complémentaire, explicite ou implicite, sur le rapport des variabilités propres des deux grandeurs comparées.

---

[1]Il ne saurait être question d'aborder ici une étude comparative des caractéristiques théoriques de la méthode des composantes principales et de l'analyse factorielle, pour laquelle je renverrai à des ouvrages tels que ceux de Thompson (*The factorial analysis of Human Ability*) ou de Burt (*Factors of the Mind*). Rappelons seulement que les techniques de calcul employées dans les deux cas sont les mêmes, à cela près que pour aboutir aux composantes principales, il faut que les éléments placés dans la diagonale principale de la matrice des corrélations soient tous égaux à l'unité, tandis que, pour aboutir aux facteurs, il faut que ces éléments diagonaux, les communautés, soient estimés par approximations successives. La trace de la matrice, somme des éléments diagonaux, a, dans le premier cas, sa valeur maxima n et, dans le second, la valeur minima compatible avec une solution réelle du problème posé, les deux techniques correspondant ainsi, en quelque sorte, à deux interprétations extrêmes des faits observés.

Pour $n = 3$, les trois droites sont généralement distinctes. La solution du système des trois équations de définition de T est donnée par : $r_{T1}^2 = r_{12}r_{13}/r_{23}$, $r_{T2}^2 = r_{12}r_{23}/r_{13}$, $r_{T3}^2 = r_{13}r_{23}/r_{12}$. Si ces saturations sont inférieures à 1, c'est-à-dire si les trois coefficients de corrélation partielle $r_{12\cdot3}$, $r_{13\cdot2}$, $r_{23\cdot1}$, sont positifs, la solution ainsi obtenue est acceptable et le facteur général suffit à expliquer les corrélations observées. Si l'une ou l'autre des saturations est plus grande que 1 (cas Heywood), deux facteurs sont nécessaires et la solution est indéterminée. Dans le premier cas, le plus fréquent de beaucoup, (D'') peut s'écrire:

$$\frac{X_1 - \overline{X}_1}{\sigma_1} r_{23} = \frac{X_2 - \overline{X}_2}{\sigma_2} r_{13} = \frac{X_3 - \overline{X}_3}{\sigma_3} r_{12}$$

équations remarquables en ce qu'elles donnent pour la relation existant entre X_1 et X_2 une expression qui ne dépend pas de r_{12} mais bien de r_{13} et r_{23} : l'introduction d'une troisième variable lève ainsi l'indétermination qui pesait sur la relation existant entre les deux premières. Les lignes de régression de X_2 en X_1 et de X_1 en X_2 correspondent respectivement aux cas où les coefficients de corrélation partielle $r_{23\cdot1}$ et $r_{13\cdot2}$ sont nuls; les trois droites (D), (D') et (D'') ont même projection sur le plan X_1X_2 lorsque $r_{13} = r_{23}$. Elles se superposent lorsque les trois coefficients de corrélation sont égaux.

Ce dernier résultat se généralise immédiatement, pour d'évidentes raisons de symétrie, au cas d'un nombre quelconque de variables: lorsque les coefficients de corrélation de tous les X pris deux à deux sont égaux, et dans cette hypothèse seulement, les droites (D') et (D'') se superposent à (D). Elles ne coïncident d'ailleurs pas point par point, la valeur estimée du facteur général t relatif à un individu n'étant pas identique à la valeur calculée correspondante f de la première composante principale.

L'équation caractéristique de la matrice des corrélations est facile à résoudre dans ce cas. La racine correspondant au plus grand axe est égale à $1 + (n - 1)r$, r étant le coefficient de corrélation commun de toutes les variables; les $(n - 1)$ autres racines, correspondant à autant d'axes secondaires de même longueur et de direction indéterminée, sont égales à $(1 - r)$, quel que soit n; les ellipsoïdes d'égale probabilité sont donc d'autant plus allongés, pour une même valeur de r, que n est plus grand. Le variance moyenne qui subsiste, une fois que la première composante est fixée, et qui est imputable à l'ensemble des $(n - 1)$ composantes négligées est égale à $(1 - r)(n - 1)/n$; la valeur estimée

du coefficient de corrélation est, dans les mêmes conditions, $r + (1 - r)/n$.
Les équations de la droite (D′) sont:

$$\frac{X_1 - \overline{X}_1}{\sigma_1} = \frac{X_2 - \overline{X}_2}{\sigma_2} = \cdots = \frac{X_n - \overline{X}_n}{\sigma_n} = \sqrt{\frac{1 + (n-1)r}{n}}\; t.$$

La valeur de F correspondant à un individu dont les mesures sont
x_1, $x_2 \cdots x_n$ est:

$$f = \sqrt{\frac{1}{n[1 + (n-1)r]}}\; \sum \frac{x_i - \overline{X}_i}{\sigma_i}\; ,$$

ce qui donne pour les valeurs théoriques x_1', $x_2' \cdots x_n'$ correspondant
aux valeurs observées x_1, x_2, $\cdots x_n$:

$$\frac{x_1' - \overline{X}_1}{\sigma_1} = \frac{x_2' - \overline{X}_2}{\sigma_2} = \cdots = \frac{x_n' - \overline{X}_n}{\sigma_n} = \frac{1}{n} \sum \frac{x_i - \overline{X}_i}{\sigma_i}.$$

L'analyse factorielle de la même matrice est complète après extraction du facteur général. La variance résiduelle moyenne imputable aux facteurs spécifiques est $(1 - r)$ quel que soit n; les coefficients de corrélation sont exactement restitués. Les équations de (D″) ont la même forme que celles de (D′), T, estimation du facteur général y remplaçant F, mesure de la première composante principale. La valeur de T correspondant à l'individu x_1, $x_2 \cdots x_n$ est estimée par régression à:

$$t = \frac{\sqrt{r}}{1 + (n-1)r} \sum \frac{x_i - \overline{X}_i}{\sigma_i}\; ,$$

ce qui donne pour x_1'', $x_2'' \cdots x_n''$, valeurs calculées par le facteur général pour x_1, x_2, $\cdots x_n$:

$$\frac{x_1'' - \overline{X}_1}{\sigma_1} = \frac{x_2'' - \overline{X}_2}{\sigma_2} = \cdots$$

$$= \frac{x_n'' - \overline{X}_n}{\sigma_n} = \sqrt{\frac{nr}{1 + (n-1)r}}\; \frac{1}{n} \sum \frac{x_i - \overline{X}_1}{\sigma_i}.$$

On voit que T diffère de F, comme x_1'', x_2'', $\cdots x_n''$ diffèrent de
x_1', x_2', $\cdots x_n'$, ce qui explique que les résultats d'une estimation de la variance liée, ou du calcul d'un coefficient de corrélation restitué, ne soient pas les mêmes avec les deux méthodes. D'une façon plus précise:

$$\frac{T}{F} = \frac{x_1'' - \overline{X}_1}{x_1' - \overline{X}_1} = \frac{x_2'' - \overline{X}_2}{x_2' - \overline{X}_2} = \cdots \frac{x_n'' - \overline{X}_n}{x_n' - \overline{X}_n} = \sqrt{\frac{nr}{1 + (n-1)r}}\; ,$$

les rapports étant d'autant plus proches de 1 que r est lui-même plus voisin de 1 ou que n est plus grand. La différence entre les valeurs estimées de la variance résiduelle moyenne calculées par la première composante principale d'une part et par le facteur général d'autre part, est $- (1 - r)/n$, l'estimation fournie par l'analyse factorielle étant $1 - r$. La différence correspondante pour les coefficients de corrélation est $- (1 - r)/n$, la valeur exacte donnée par l'analyse factorielle étant r. Les estimations fournies par les deux méthodes convergent donc lorsque r tend vers 1 ou lorsque n croît. Si, r restant fixé, n augmente indéfiniment, la valeur limite est celle qu'avait donnée d'emblée l'analyse factorielle.

Il existe un dernier cas particulier important, dont nous avons d'ailleurs déjà fait mention en définissant la droite (D''). C'est celui où les coefficients de corrélation peuvent tous être mis sous la forme d'un produit de saturations $r_{pq} = r_{T_p} r_{T_q}$, où, autrement dit, la matrice des corrélations est strictement hiérarchique. Dans ce cas, les trois droites sont distinctes et ont chacune leur signification propre. (D) qui ne dépend pas des coefficients de corrélation, peut cependant être considérée comme définissant la relation qui unirait les n variables si, les distributions marginales restant inchangées, les coefficients de corrélation r_{pq} devenaient tous égaux à leur valeur moyenne r. L'analyse factorielle étant complète après extraction du premier facteur, les coefficients de corrélation sont exactement restitués, aux erreurs d'échantillonnage près et la droite (D'') peut être considérée comme rassemblant le maximum d'information sur la structure de la distribution totale. Cette droite présente par ailleurs un caractère d'invariance, qu'elle partage avec (D) et que ne possède pas (D'). Si l'on ajoute en effet à la série des mesures déjà faites, celles d'une nouvelle grandeur X_s et si celle-ci peut être caractérisée par une saturation r_{T_s} telle que pour tout X_p on ait $r_{ps} = r_{T_p} r_{T_s}$, les équations de la nouvelle droite (D'') s'obtiendraient simplement en complétant la série des égalités déjà écrites par un nouveau terme $(X_s - \overline{X}_s)/r_{T_s} \sigma_s$. Dans les mêmes hypothèses, la droite (D') devra être entièrement recalculée, mais il est facile de voir que (D') diffère d'autant moins de (D'') que le nombre des variables est plus grand.

Si $\bar{a}^2 = \sum r_{T_p}^2/n$ est la variance moyenne imputable au facteur général, et si l'on suppose les X rangés dans l'ordre décroissant des saturations, la plus grande racine de l'équation caractéristique de la matrice des corrélations est, en effet, comprise entre $1 + n\bar{a}^2 - r_{T_1}^2$ et $1 + n\bar{a}^2 - r_{T_n}^2$, et diffère peu par conséquent de $1 + (n - 1)\bar{a}^2$ ou, ce qui nos hypothèses revient pratiquement au même, de $1 + (n - 1)\bar{r}$. Les $(n - 1)$

autres racines diffèrent les unes des autres et s'intercalent entre les variances résiduelles rangées par valeurs croissantes $(1 - r_{T_1}^2)$, $(1 - r_{T_2}^2)$, $(1 - r_{T_3}^2) \cdots (1 - r_{T_n}^2)$. Elles varient peu avec n et les ellipsoïdes d'égale probabilité sont d'autant plus allongés que les variables sont plus nombreuses. Comme dans le cas précédent, la variance qui subsiste une fois fixée la première composante principale tend vers $1 - \bar{a}^2$ lorsque augmente indéfiniment. Dans la même hypothèse, les saturations et coefficients de corrélation calculés à partir de la première composante principale tendront également vers les valeurs estimées par le facteur général.

A tous les cas particuliers qui viennent d'être envisagés peut être opposé le cas général où un seul facteur ne suffit pas à rendre compte des coefficients de corrélation observés et où ceux-ci ne peuvant être restitués exactement que par le jeu combiné du facteur général et de un ou plusieurs facteurs bipolaires. Mais, si l'analyse est conduite par des techniques dérivant de la méthode des moindres carrés, comme celle qu'a codifiée Burt, l'estimation du facteur général est telle que la valeur moyenne du coefficient de corrélation calculée d'après ce seul facteur est égale à la valeur moyenne des coefficients de corrélation observés. La variance résiduelle moyenne $1 - \bar{a}^2$ est donc proche de $1 - \bar{r}$.

Il n'est plus possible, par ailleurs, de donner une expression générale de la plus grande racine de l'équation caractéristique, mais une formule connue en donne une valeur approchée qui, dans notre notation s'écrit $1 + (n - 1)\bar{r}$, valeur identique à celle que nous avons obtenue par un calcul exact dans le cas de l'intercorrélation constante et retrouvée comme approximation pur une matrice hiérarchique. Comme précédemment, la variance qui subsiste en moyenne une fois fixée la première composante principale peut être estimée à $(1 - \bar{r})(n - 1)/n$ et tend vers $1 - \bar{r}$ si $\bar{r}$ restant constant, n croît indéfiniment. Il xy a donc là encore convergence de (D') vers (D'').

Nous sommes maintenant en mesure de répondre à la question que nous nous étions posée. La droite (D) ne peut être retenue que si les différents coefficients de corrélation peuvent légitimement être considérés comme égaux, circonstance évidemment exceptionnelle. Les droites (D') et (D'') sont utilisables dans tous les cas et, puisque (D') tend vers (D'') et s'en rapproche d'autant plus que, les mesures étant multipliées, nous disposons de plus d'informations sur notre matériel, (D'') doit être préférée à (D'). S'il se trouve qu'un facteur général puisse suffire à une restitution exacte, aux erreurs d'échantillonnage près, des coefficients de corrélation observés, la droite (D'') traduira

complètement les relations existant entre les n variables, les écarts entre X calculés et X observés étant indépendants les uns des autres et devant être tenus pour fortuits. S'il n'en est pas ainsi et si, une fois extrait le facteur général, il subsiste des corrélations significatives, on doit conclure que la droite (D″) n'épuise pas les informations que nous apportent les données sur les relations unissant les variables étudiées, mais la fraction qu'elle en extrait est d'autant plus grande que $\bar{r}$ est plus proche de 1.

On remarquera que la solution qui est proposée fait disparaître la difficulté, fréquente dans la pratique biométrique, du choix d'une grandeur de référence. Celle qui intervient dans nos calculs, F ou T, suivant qu'il s'agit de la première composante principale et de (D′), ou du facteur général et de (D″), est une moyenne pondérée de tous les $(X - \overline{X})/\sigma$, relatifs à un individu. Dans le cas où l'on croit pouvoir utiliser (D), le paramètre correspondant serait simplement la moyenne s des écarts rèduits des X.

Il ne sera peut-être pas inutile d'illustrer l'étude qui précède par un exemple concret. Celui-ci a été pris dans un mémoire en cours d'impression dans les *Archives de Zoologie expérimentale et générale*, et porte sur un Crabe Oxyrhynque, *Maïa squinado*, sur lequel ont été mesurées les dimensions principales du céphalothorax et de différents appendices. Les X sont les logarithmes népériens P_0 , M_0 , M_1 , M_2 , M_3 , M_4 et L, des huit mesures pratiquées sur chacun des 301 animaux étudiés. Ces X ont des distributions sensiblement normales et leurs coefficients de corrélation, très élevés, s'écartent peu de leur valeur moynne $\bar{r} = 0{,}9752$, comme le montre le Tableau I. Le Tableau II donne les saturations en F et en T, r_{FX} et r_{TX} , calculées par sommation et itération à partir du Tableau I, et les coefficients directeurs des droites (D) et (D′) et (D″), soit respectivement, σ_X , $r_{FX}\sigma_X$ et $r_{TX}\sigma_X$ [1]. Les trois solutions sont manifestement très voisines l'une de l'autre, fait en relation évidente avec la faible dispersion des coefficients de corrélation.

La comparaison entre les résultats des différentes techniques de calcul est plus instructive lorsqu'elle porte sur les relations unissant deux variables X_1 et X_2 . Si G est la grandeur de référence prise par convention comme variable indépendante, on est toujours en droit d'écrire:

$$X_2 - \overline{X}_2 = \frac{r_{G1}}{r_{G2}}\frac{\sigma_1}{\sigma_2}(X_1 - \overline{X}_1)$$

---

[1] Le facteur général étant seul recherché, les communautés utilisées dans cette analyse factorielle ont reçu, par approximations successives, la plus petite valeur possible. Les saturations en T, obtenues par une analyse factorielle complète diffèrent d'ailleurs très peu de celles qui sont données ici. On les trouvera dans mon mémoire des *Archives de Zoologie expérimentale et générale*.

TABLEAU I.

	P_0	C_0	M_0	M_1	M_2	M_3	M_4	L
P_0	. . . . .	,9899	,9892	,9742	,9703	,9702	,9616	,9592
C_0	,9899	. . . . .	,9929	,9776	,9735	,9722	,9669	,9611
M_0	,9892	,9929	. . . . .	,9798	,9757	,9742	,9689	,9658
M_1	,9742	,9776	,9798	. . . . .	,9900	,9896	,9832	,9672
M_2	,9703	,9735	,9757	,9900	. . . . .	,9911	,9861	,9674
M_3	,9702	,9722	,9742	,9896	,9911	. . . . .	,9861	,9660
M_4	,9616	,9669	,9689	,9832	,9861	,9861	. . . . .	,9592
L	,9592	,9611	,9658	,9672	,9674	,9660	,9592	. . . . .

TABLEAU II.

	X	σ_X	r_{FX}	$r_{FX}\sigma_X$	r_{TX}	$r_{TX}\sigma_X$
		(D)		(D')		(D'')
P_0	4,8804	0,2008	0,9875	0,1983	0,9854	0,1979
C_0	4,1951	0,1789	0,9900	0,1781	0,9887	0,1769
M_0	4,3386	0,1749	0,9916	0,1734	0,9908	0,1733
M_1	4,4892	0,1440	0,9935	0,1431	0,9934	0,1430
M_2	4,3836	0,1361	0,9925	0,1351	0,9921	0,1350
M_3	4,2413	0,1332	0,9920	0,1321	0,9913	0,1320
M_4	4,0497	0,1269	0,9872	0,1253	0,9850	0,1250
L	5,1909	0,1036	0,9788	0,1014	0,9739	0,1009

TABLEAU III.

	$\bar{X}$	σ_X	r_{FX}	$r_{FX}\sigma_X$	r_{TX}	$r_{TX}\sigma_X$
		(D)		(D')		(D'')
P_0	4,8804	0,2008	0,9867	0,1981	0,9801	0,1968
M_1	4,4892	0,1440	0,9936	0,1431	0,9944	0,1432
M_2	4,3836	0,1361	0,9927	0,1351	0,9925	0,1351
L	5,1909	0,1036	0,9840	0,1019	0,9754	0,1010

On obtient (D) en supposant que l'on peut admettre l'égalité de r_{1G} et r_{2G} . (D') et (D'') correspondent respectivement à $G = F$ et $G = T$, la grandeur de référence pouvant être définie par un nombre plus ou moins grand de variables. Si $n = \xi$ et si l'on ne se trouve pas dans le cas Heywood, T est déterminé exactement par l'adjonction d'une variable auxiliaire X_3 au couple de variables principales X_1 et X_2 et la solution factorelle correspond à $G = X_3$. Les deux droites de régres-

sion de X_2 en X_1 et de X_1 en X_2 correspondent respectivement à $G = X_3 = X_1$ et $G = X_3 = X_2$.

Pour appliquer ces formules au cas des *Maïa squinado*, quatre des huit grandeurs étudiées P_0, M_1, M_2 et L ont été choisies et les co-efficients angulaires des six droites, désignées ici par (P_0, L), (M_1, L), (M_2, L); (P_0, M_2), (M_1, M_2) et (P_0, M_1), ont été calculés par différents procédés et à partir d'une, deux ou six variables auxiliaires. Le tableau III correspond au Tableau II, mais comporte seulement quatre lignes: il a été calculé à partir des six coefficients de corrélation existant entre les quatre variables étudiées, par les mêmes méthodes que celles qui ont servi à l'analyse de l'ensemble des résultats.

Le Tableau IV donne les pentes des six droites calculées par trois méthodes. Les huit premières lignes donnent les résultats obtenus en prenant successivement comme variable auxiliaire X_3 une des grandeurs P_0, C_0, $\cdots$ ou L, et en appliquant, soit la technique des composantes principales (D_3'), soit la technique factorielle (D_3''). Les nombres en italiques correspondent à $X_3 = X_1$ et $X_3 = X_2$; ils donnent la pente de (D) dans la colonne (D_3') et les pentes des deux lignes de régression classiques dans la colonne (D_3''). La neuvième ligne donne la pente de la droite (D) obtenue en négligeant les différences existant entre les co-efficients de corrélation; elle est accompagnée de son écart-type qui est intermédiaire entre les écarts-types des deux coefficients de régression correspondants. Les lignes (D_4') et (D_4'') ont été calculées à partir des données du Tableau III, et les lignes (D_8') et (D_8'') à partir du Tableau II. Elles donnent les valeurs numériques correspondant aux relations obtenues par la technique des composantes principales (D') et par la technique factorelle (D'') pour quatre variables P_0, M_1, M_2 et L, et pour les huit X.

Pour un couple donné de variables X_1, X_2, (D) est, par définition, indépendant de toute corrélation et les coefficients de régression ne dépendent que de r_{12}. Toutes les autres estimations de la pente de la droite (X_1, X_2) changent avec la nature et le nombre des variables auxiliaires choisies et dépendent aussi de la méthode de calcul adoptée. Les pentes estimées sont toujours comprises entre celles qui correspondent aux deux lignes de régression, et assez éloignées de l'une et de l'autre. On pourra constater que les pentes des (D_4'') et des (D_8'') sont très approximativement égales aux moyennes des pentes des (D_3'') correspondantes, droites de régression non comprises, et que les pentes des (D_4') et (D_8') sont proches des moyennes des pentes des (D_3''), droites de régression comprises et non pas des moyennes des pentes des (D_3'). Une restitution beaucoup plus exacte des pentes des (D') est d'ailleurs obtenue en remplaçant dans la moyenne des pentes des (D_3'') les pentes

TABLEAU IV.

	(P_0, L)		(M_1, L)		(M_2, L)		(P_0, M_2)		(M_1, M_2)		(P_0, M_1)	
	(D'_3)	(D''_3)	(D'_3)	(D''_3)	(D'_3)	(D''_3)	(D'_3)	(D''_3)	(D'_3)	(D''_3)	(D'_3)	(D''_3)
P_0	1,938	2,021	1,397	1,412	1,319	1,329	1,475	1,520	1,059	1,062	1,394	1,431
C_0	1,958	1,996	1,398	1,414	1,319	1,331	1,484	1,500	1,060	1,063	1,400	1,412
M_0	1,954	1,985	1,397	1,410	1,318	1,327	1,482	1,496	1,060	1,063	1,399	1,408
M_1	1,943	1,952	1,390	1,437	1,324	1,345	1,468	1,452	1,058	1,069	1,394	1,359
M_2	1,940	1,944	1,401	1,423	1,314	1,358	1,475	1,432	1,058	1,048	1,385	1,367
M_3	1,941	1,947	1,401	1,424	1,325	1,348	1,465	1,444	1,057	1,056	1,385	1,367
M_4	1,940	1,943	1,402	1,425	1,326	1,350	1,463	1,439	1,057	1,055	1,384	1,364
L	1,938	1,859	1,390	1,344	1,314	1,271	1,471	1,463	1,058	1,058	1,391	1,383
(D)	1,9382 ∓ 0,0816		1,3900 ∓ 0,0204		1,3187 ∓ 0,0192		1,4754 ∓ 0,0206		1,0580 ∓ 0,0086		1,3944 ∓ 0,0182	
(D'_4)	1,9433		1,4036		1,3253		1,4664		1,0590		1,3846	
(D''_4)	1,9475		1,4171		1,3367		1,4569		1,0600		1,3743	
(D'_8)	1,9555		1,4108		1,3321		1,4680		1,0591		1,3861	
(D''_8)	1,9612		1,4178		1,3382		1,4655		1,0594		1,3833	

des deux lignes de régression par le double de la pente de (D) corre-
spondante. Par là s'explique bien que (D') soit toujours compris
entre (D) et (D'') et que, confondu avec (D) lorsqu'aucune variable
auxiliaire n'est utilisée, il s'en détache pour $n = 3$ et se rapproche
graduellement de (D''), lorsque le nombre des variables auxiliaires,
passe de 1 à 2, puis de 2 à 6.

Ainsi se vérifient les propositions que nous avions énoncées et se
justifie la préférence que nous avions donnée à la droite d'interpolation
obtenue par analyse factorielle.

REFERENCES

Burt (C.)—Factors of the Mind (*London University Press*, 1940).

Cramer (M.)—Mathematical methods of Statistics (*New-York Princeton University Press*, 1946.)

Darmois (G.)—Sur la détermination de l'axe d'un nuage rectiligne de points (*Biometrics*, 1954, *10*, no. 1, 180).

Kermack (K. A.) et Haldane (J. B. S.)—Organic correlation and Allometry (*Biometrika*, 1950, *37*, 30).

Kruskal (W. H.)—On the uniqueness of the line of organic correlation (*Biometrics*, 1953, *9*, no. 1, 47).

Quenouille (M. H.)—Associated Measurements (*London, Butterworths Sc. Pub.* 1952).

Teissier (G.)—La relation d'allométrie; sa signification statistique et biologique) (*Biometrics*, 1948, *4*, no. 1, 14).

Teissier (G.)—Allométrie de taille et variabilité chez *Maia squinado* (*Arc.h Zool. exp. et gen.*, 1955, *92*, 141.

Thompson (G. M.)—The factorial analysis of human ability (*London University Press*, 1948).

THE VARIANCE OF THE GENETIC CORRELATION COEFFICIENT

E. C. R. REEVE*

Institute of Animal Genetics, Edinburgh

1. *Introduction*

A genetic correlation coefficient measures the degree of association between the genetic variations of two quantitative characters in a given population, e.g. wing and thorax length in *Drosophila* (Reeve and Robertson, 1953), or weight and market score in pigs (Hazel, 1943). Genetic correlations are frequently calculated in work on quantitative inheritance and animal breeding research, but no method is available for assessing their accuracy when they are estimated from a given body of data. It is the purpose of this paper to develop formulae for the large-sample variance of a genetic correlation when it is estimated in various ways from parent-offspring covariances or correlations. Qualifications to be borne in mind when using these formulae will be discussed later.

Consider a progeny test in which a number of pair-matings are made of parents selected at random, and two characters are measured on both parents and progeny from each mating. Let a and x be the parent phenotypes and b and y the corresponding progeny (or progeny mean) phenotypes for the two characters, so that (a, b) refer to character 1 and (x, y) to character 2, in the two generations. Then the genetic correlation (r_G) between characters 1 and 2 may be estimated in two ways, both of which appear to have been first suggested by Hazel (1943). The alternative formulae are as follows:

$$r_G = \left[\frac{r_{ay} \cdot r_{xb}}{r_{ab} \cdot r_{xy}} \right]^{1/2} = \left[\frac{\{ay\} \cdot \{xb\}}{\{ab\} \cdot \{xy\}} \right]^{1/2} \tag{1a}$$

$$r_G = \frac{\{ay\} + \{xb\}}{2[\{ab\} \cdot \{xy\}]^{1/2}} \tag{1b}$$

where $\{ay\}$ is the covariance between a and y, etc., and may in practice be replaced by the sum of products $\sum (a - \bar{a})(y - \bar{y})$, provided that all four covariances have the same number of degrees of freedom.

---

*Member of the Agricultural Research Council's scientific staff.

It may be easily shown that both formulae give estimates of the same parameter; for suppose that H_1 and H_2 are the genetic values of the two characters in an individual (i.e. the expected phenotypes of individuals of the same genotype, in a given environment), then, assuming that non-additive genetic variations can be ignored, and writing E for 'Expected value of',

$$
\left.
\begin{aligned}
E\{ay\} &= E\{xb\} = \tfrac{1}{2}\operatorname{cov}\{H_1, H_2\} \\
E\{ab\} &= \tfrac{1}{2}\sigma^2\{H_1\}, \qquad E\{xy\} = \tfrac{1}{2}\sigma^2\{H_2\} \\
r_G &= \operatorname{cov}\{H_1, H_2\}/\sigma\{H_1\} \cdot \sigma\{H_2\}
\end{aligned}
\right\}
\tag{2}
$$

Formulae (1a) and (1b) follow immediately, and either may be derived as a least squares estimate, after taking logarithms in the third equation of (2) to give a linear formula for r_G, depending on whether we combine $\{ay\}$ and $\{xb\}$ first, to give a separate estimate of cov $\{H_1, H_2\}$ (formula 1b), or carry out the whole Least Squares operation in one step (formula 1a).

As we shall see, both formulae give estimates with the same variance in large samples, but obviously (1b) must always give a smaller estimate than (1a), except when $\{ay\} = \{xb\}$ and the two estimates are equal. Probably both formulae lead to biassed estimates in finite samples, but I have been unable to calculate the extent of this bias or to suggest which estimate will generally be least biassed. Formula (1b) has the advantage of giving a non-imaginary estimate when either $\{ay\}$ or $\{xb\}$ comes out negative and the other covariances are positive, and for this reason this formula tends to be preferred (Morley, 1951). The question of their relative merits obviously needs further study.

The four variables may be individual phenotypes of parent and offspring, but a more accurate estimate is obtained when a and x are mid-parent values and b and y are the means of, say, n progeny of a family. This will apply to a test in which pair matings are made and a family of full-sib progeny from each mating is measured; but alternatively we may have measurements on families of half-sib progeny and only on the parent common to each family. In this case a and x are single parent phenotypes and b and y are means of n half-sibs. Formulae (1) and (2) apply in both cases, and they will be treated together. Modifications when the characters are sex-limited, or selection and assortative mating of parents are practised, will be discussed later. Sex-linked effects will be ignored (the separation of sex-linked and autosomal genetic variances for one character have been discussed by Reeve (1953), but the separation of genetic correlations of the two types would raise too many complications for treatment here).

2. *Variance when parents are a random sample.*

For convenience, we shall deal with formula (1a), and will show later that (1b) has the same variance. Suppose the four covariances are all estimated from a single progeny test, consisting of f families of n progeny and their corresponding parents. We then have a quadri-variate distribution of a, b, x and y, each of which is correlated with the others, so that errors in the four covariances of (1a) will be correlated.

Taking logarithms in (1a), differentiating, squaring and taking expectations, we obtain approximately:

$$V(r_G) = \frac{r_G^2}{4} \left[\frac{V\{ay\}}{\{ay\}^2} + \frac{V\{xb\}}{\{xb\}^2} + \frac{V\{ab\}}{\{ab\}^2} + \frac{V\{xy\}}{\{xy\}^2} \right.$$

$$+ \frac{2\,\mathrm{Cov}\,[\{ay\},\{xb\}]}{\{ay\}\cdot\{xb\}} + \frac{2\,\mathrm{Cov}\,[\{ab\},\{xy\}]}{\{ab\}\cdot\{xy\}}$$

$$- \frac{2\,\mathrm{Cov}\,[\{ay\},\{xy\}]}{\{ay\}\cdot\{xy\}} - \frac{2\,\mathrm{Cov}\,[\{ab\},\{xb\}]}{\{ab\}\cdot\{xb\}}$$

$$\left. - \frac{2\,\mathrm{Cov}\,[\{ab\},\{ay\}]}{\{ab\}\cdot\{ay\}} - \frac{2\,\mathrm{Cov}\,[\{xb\},\{xy\}]}{\{xb\}\cdot\{xy\}} \right] \qquad (3)$$

where V stands for "Variance of", and $\mathrm{Cov}\,[\{ay\},\{xb\}]$ stands for the sampling covariance of the covariances of a with y and x with b. To make further progress, we must assume that the four variates are normally distributed. Then we have, approximately, (writing ρ for a population correlation coefficient):

$$V\{ay\} = \frac{1}{f}\,\sigma_a^2\sigma_y^2(1 + \rho_{ay}^2),$$

so that

$$\frac{V\{ay\}}{\{ay\}^2} = \frac{1}{f}\left(1 + \frac{1}{\rho_{ay}^2}\right), \qquad \text{etc.} \qquad (4)$$

Also

$$\mathrm{Cov}\,[\{ab\},\{xy\}] = E\left[\left\{\frac{\sum ab - f\bar{a}\bar{b}}{f-1} - \mu_{ab}\right\}\left\{\frac{\sum xy - f\bar{x}\bar{y}}{f-1} - \mu_{xy}\right\}\right]$$

$$= \frac{1}{f}\,[\mu_{abxy} - \mu_{ab}\mu_{xy}] \qquad (5)$$

ignoring terms of order $1/f^2$.

Similarly

$$\text{Cov}\;[\{ab\},\,\{ay\}] = \frac{1}{f}\,[\mu_{a^2by} - \mu_{ab}\mu_{ay}] \tag{6}$$

The moments μ_{abxy} etc. may be found by extending the M.G.F. for a bivariate population:

$$M(t_a,\,t_b,\,t_x,\,t_y) = \exp\,[\tfrac{1}{2}(\textstyle\sum t_a^2\sigma_a^2 + 2\sum \rho_{ab}\sigma_a\sigma_b t_a t_b)]$$

We obtain

$$\mu_{abxy} = \sigma_a\sigma_b\sigma_x\sigma_y[\rho_{ab}\rho_{xy} + \rho_{ax}\rho_{by} + \rho_{ay}\rho_{bx}] \tag{7}$$

and

$$\mu_{a^2bx} = \sigma_a^2\sigma_b\sigma_x[\rho_{bx} + 2\rho_{ab}\cdot\rho_{ax}] \tag{8}$$

Then from (5) and (7)

$$\frac{\text{Cov}\;[\{ab\},\,\{xy\}]}{\{ab\}\cdot\{xy\}} = \frac{1}{f}\,\frac{\rho_{ax}\cdot\rho_{by} + \rho_{ay}\cdot\rho_{bx}}{\rho_{ab}\cdot\rho_{xy}} \tag{9}$$

and from (6) and (8)

$$\frac{\text{Cov}\;[\{ab\},\,\{ax\}]}{\{ab\}\cdot\{ax\}} = \frac{1}{f}\left[1 + \frac{\rho_{bx}}{\rho_{ab}\cdot\rho_{ax}}\right] \tag{10}$$

Substituting formulae of type (4), (9) and (10) in (3) we obtain the rather impressive formula:

$$V(r_G) = \frac{r_G^2}{4f}\left[-4 + \frac{1}{[ay]^2} + \frac{1}{[xb]^2} + \frac{1}{[ab]^2} + \frac{1}{[xy]^2}\right.$$

$$+\,2\,\frac{[ax]\cdot[by] + [ay]\cdot[xb]}{[ab]\cdot[xy]} + 2\,\frac{[ax]\cdot[by] + [ab]\cdot[xy]}{[ay]\cdot[xb]}$$

$$\left.-\,2\,\frac{[ax]}{[ay]\cdot[xy]} - 2\,\frac{[ax]}{[ab]\cdot[xb]} - 2\,\frac{[by]}{[ab]\cdot[ay]} - 2\,\frac{[by]}{[xb]\cdot[xy]}\right] \tag{11}$$

where square brackets, in order to save subscripts, indicate correlation coefficients, e.g. $[ax] = \rho_{ax}$, and should not be confused with $\{ax\} =$ the covariance between a and x.

The correlation coefficients in (11) may be expressed in terms of the so-called heritabilities of the two characters and their genetic and phenotypic correlations. Suppose a and x are the mid-parent (or common parent) phenotypes, and b and y are the means of n progeny, where n is the same for all matings. Using subscripts 1 for character (a, b) and 2 for character (x, y) and writing h^2 for the heritability of individuals—i.e. the fraction of the phenotypic variance which is genetic—we have the following relations, which hold good for progeny

tests in which either the progeny are full sibs and mid-parent values are used, or the progeny are $\frac{1}{2}$-sibs and the one common parent is measured:

$$[ab] = h_1^2 \Big/ \left(\frac{2Bk}{n}\right)^{1/2}$$

$$[xy] = h_2^2 \Big/ \left(\frac{2Yk}{n}\right)^{1/2}$$

$$[xb] = h_1 h_2 r_G \Big/ \left(\frac{2Bk}{n}\right)^{1/2}$$

$$[ay] = h_1 h_2 r_G \Big/ \left(\frac{2Yk}{n}\right)^{1/2}$$

$$[ax] = r_P$$

$$[by] = \left(r_P + \frac{n-1}{2k} h_1 h_2 r_G\right) \Big/ (BY)^{1/2}$$

$$B = 1 + \frac{n-1}{2k} h_1^2$$

$$Y = 1 + \frac{n-1}{2k} h_2^2$$

$$(12)$$

where r_P = phenotypic correlation between characters 1 and 2 for individuals, $k = 1$ in a test using mid-parent values and families of full sibs, and $k = 2$ for a test using one parent and families of $\frac{1}{2}$ sibs.

These formulae are easily derived. Thus, for mid-parent and full-sib families, $[ab] = h_1^2 \sigma_a / \sigma_b$, $\sigma_a^2 = \frac{1}{2}\sigma_1^2$, and $\sigma_b^2 = 1/n \, [1 + (n-1)r_{oo}]\sigma_1^2$, where σ_1^2 is the phenotypic variance of character (a, b) in individuals, and $r_{oo} = \frac{1}{2}h_1^2$ is the correlation between full sibs (cf. Appendix and Reeve, 1953).

Substituting equations (12) in (11) and simplifying, we obtain:

$$V(r_G) = \frac{1}{f}\left[\frac{1}{2}(1 - r_G^2)^2 + \frac{1}{2}(1 - r_G^2)\left(\frac{1}{D} - \frac{r_P r_G}{C}\right) + \frac{2k}{n}\left(\frac{r_G}{D} - \frac{r_P}{C}\right)^2 \right.$$

$$\left. + \frac{(1 - r_G^2)}{n}\left\{\frac{k(1 - r_P^2)}{C^2} - \frac{1}{2D} + \frac{r_P r_G}{2C}\right\}\right] \qquad (13)$$

where $C = h_1 h_2$, $1/D = \frac{1}{2}(1/h_1^2 + 1/h_2^2)$, and n progeny are measured from each of f families. As a further variation, if we are dealing with sex-limited characters, in a test such that one parent and families of full sibs of the same sex are measured, then $k = 1$ and all items in (13) are multiplied by 2 except the first term: $(1 - r_G^2)^2/2f$.

It follows from (13) that, if a given total number of individuals are to be measured (i.e. nf is constant), the variance of r_G is least when $n = 1$, and (13) then reduces to:

$$V(r_G) = \frac{1}{f}\left[\frac{1}{2}(1 - r_G^2)^2 + \frac{k(1 - r_G^2)(1 - r_P^2)}{C^2} + 2k\left(\frac{r_G}{D} - \frac{r_P}{C}\right)^2\right] \qquad (14)$$

But in practice, of course, it is often easier to increase n than f.

In tests with poultry and livestock, usually several sires are each mated to a number of females, and there may be several progeny from each mating, so that each sire provides families of full sibs in a population of half-sibs. Within each sire group we may take half the female parent's phenotypes as the mid-parent values (since the sire's phenotypes for the two characters are constant, whether we can measure them or not, and may be scored as zero). Equations (12) then apply with $k = 1$ and with $\frac{1}{2}n$ always substituted for n. We may thus estimate the genetic correlation by pooling the four covariances within sire groups and using formula (1a) or (1b), and formula (13) will then give an approximate estimate of its variance if we take f as the total number of female parents minus the number of sires and n as $\frac{1}{2}$ the average number of progeny per female parent (the method of averaging n will be discussed in a later section).

Formulae (13) and (14) are expressed in terms of the population values of r_G, r_P, C and D, which will not normally be known; and the best we can do is to use estimates either from the progeny test in question or from other sources—better estimates of r_P, C and D will often be available. Significance tests using the variance of r_G will, of course, be very unreliable, since the variance is nearly proportional to $(1 - r_G^2)$, and may be appreciably affected by a small error in estimating r_G. Moreover, the sampling distribution of r_G is probably at least as skew as that of the product-moment correlation over most of its range, so that its variance, even if known accurately, would not be very helpful in setting confidence limits, unless we have an extremely large sample and an approach to normality. Nevertheless, an approximate variance for an estimate of r_G is better than no guidance whatever as to its accuracy, and it will also enable us to calculate the size of progeny test necessary to give an estimate with a given level of accuracy. It is to be hoped that a study of the actual sampling distribution of r_G will lead to a satisfactory method of setting confidence limits to a sampling estimate.

We may note that, if r_G is zero, assuming that r_P is also small enough to be ignored (i.e. there is little or no environmental correlation between the two characters), then the sampling variance of r_G reduces to

$$V = \frac{1}{f}\left[\frac{1}{2} + \frac{1}{2D}\right] = \frac{1}{f}\left[\frac{1+D}{2D}\right] \tag{15}$$

where D, as before, is the harmonic mean of h_1^2 and h_2^2 , and will lie between 1 and 0. This formula may be compared with the variance of a product-moment correlation coefficient whose population value is 0, which is approximately $1/f$ for a sample of f pairs.

Some allowance should be made for the fact that sample estimates have to be substituted for the various statistical parameters in calculating formulae 13–15. For this purpose it seems best to take f as two less than the number of families.

3. *Effect of selection and assortative mating of parents.*

An alternative procedure will now be considered. Selection and assortative mating (i.e. picking out extreme $+$ and $-$ phenotypes and mating like with like) considerably reduces the sampling variances of the regression coefficients of progeny on parent, and does not alter the expected values of the regressions on the mid-parent value of the selected character, apart from possible bias due to non-additive genetic effects (Reeve 1953). We might, therefore, run two separate progeny tests, selecting and mating assortatively for character 1 in the first and for character 2 in the second. The regressions of b and y on a are estimated from the first test, and of b and y on x are estimated from the second. Using B to symbolise a regression coefficient, we then have:

$$r_G = \left[\frac{B(y \cdot a)}{B(b \cdot a)} \cdot \frac{B(b \cdot x)}{B(y \cdot x)}\right]^{1/2} = \left[\frac{\{ay\}_1}{\{ab\}_1} \cdot \frac{\{xb\}_2}{\{xy\}_2}\right]^{1/2} \tag{16}$$

where suffixes 1 and 2 here indicate the test supplying the estimate. Proceeding as before, since covariances from the two tests are uncorrelated, formula (11) reduces to:

$$V(r_G) = \frac{r_G^2}{2f}\left(\frac{1}{[ay]_1^2} + \frac{1}{[ab]_1^2} + \frac{1}{[xb]_2^2} + \frac{1}{[xy]_2^2}\right.$$

$$\left. - \frac{2[by]_1}{[ab]_1 \cdot [ay]_1} - \frac{2[by]_2}{[xb]_2 \cdot [xy]_2}\right) \tag{17}$$

where $f/2$ families are used in each test.

Now assume that selection plus assortative mating has the average effect in the two tests of multiplying mid-parent variance of the selected character by $(1 + L)$—which can be estimated from the tests. Dealing with the case of full-sibs and mid-parent phenotypes only, formulae

(12) must be modified as follows (some of these formulae are derived in an appendix):

$$[ax] = r_P\left(\frac{1 + L}{1 + Lr_P^2}\right)^{1/2}$$

$$[by] = \frac{r_P + \frac{1}{2}(n - 1)h_1h_2r_G + \frac{1}{2}nLh_1h_2r_GQ}{(BY)^{1/2}(1 + L)}$$

(18.1)

and here and in the formulae for $[ab]$ and $[ay]$ of (12), in test 1, we substitute: $Q = h_1^2$ and:

$$B = [1 + \tfrac{1}{2}(n - 1)h_1^2 + \tfrac{1}{2}nh_1^4L]/(1 + L)$$

$$Y = [1 + \tfrac{1}{2}(n - 1)h_2^2 + \tfrac{1}{2}nh_1^2h_2^2r_G^2L]/(1 + L)$$

(18.2)

while in calculating $[by]$, $[xy]$ and $[xb]$ in test 2, we substitute $Q = h_2^2$ and:

$$B = [1 + \tfrac{1}{2}(n - 1)h_2^2 + \tfrac{1}{2}nh_2^4L]/(1 + L)$$

$$Y = [1 + \tfrac{1}{2}(n - 1)h_1^2 + \tfrac{1}{2}nh_1^2h_2^2r_G^2L]/(1 + L)$$

(18.3)

Substituting in (17), we obtain

$$V(r_G) = \frac{2}{f(1 + L)}\left[\frac{1 - r_G^2}{2D} + \frac{1 - r_G^2}{n}\left(\frac{1}{C^2} - \frac{1}{2D}\right)\right. $$
$$\left. + \frac{2}{n}\frac{r_G}{D}\left(\frac{r_G}{D} - \frac{r_P}{C}\right)\right] \quad (19)$$

The relative value of this method depends on the magnitude of L. If the parents are not a selected sample but are mated assortatively, L is the correlation between mates, which can be made to approach unity. In an example in *Drosophila*, in which the most extreme ($+$ and $-$) half of the available parents were selected and mated assortatively, L was 1.74 (Reeve, 1953) and it should be possible to obtain values of L approaching 2, provided that the parents can be selected from a fairly large sample.

If n is fairly large (say 10 or more) we can compare the variances of the different estimates ignoring terms in $1/n$. We then have, approximately: for a single progeny test with f matings (random mating)

$$V(r_G) = \frac{1 - r_G^2}{2f}\left[\frac{1}{D} + 1 - r_G^2 - \frac{r_P r_G}{C}\right] \quad (20)$$

for two tests, each of $\frac{1}{2}f$ matings with selection and assortative mating,

$$V(r_G) = \frac{2}{1 + L}\cdot\frac{1 - r_G^2}{2Df} \quad (21)$$

D is the harmonic mean and C the geometric mean of h_1^2 and h_2^2, and $(1 - r_G^2 - r_P r_G / C)$ will generally be small compared with $1/D$. If this is the case, two tests with selection and assortative mating of parents should give an estimate of r_G with variance a little lower—in the proportion $2 : (1 + L)$—than a single test with random selection and mating, using the same number of matings.

The two tests will, of course, provide altogether 8 regressions, of which only 4 are used in (16). The 4 regressions from each test could be used to give a separate estimate, as described by Reeve (1953), and the two estimates could be averaged to obtain a final estimate of improved statistical accuracy, with a sampling variance roughly half that of (16). The difficulty about this estimate is that the amount of bias likely to arise from non-additive genetic effects is unknown, but probably larger than the bias in (16).

4. *Equality of variances for estimates* (1a) *and* (1b)

Taking the case of a single progeny test consisting of families of full- or half-sibs with their mid-parent or common parent phenotypes, which led to formula (13), we can differentiate (1b) directly, giving

$$dr_G = \frac{1}{4[\{ab\} \cdot \{xy\}]^{1/2}} \left[2\partial\{ay\} + 2\partial\{xb\} \right.$$
$$\left. - (\{ay\} + \{xb\})\left(\frac{\partial\{ab\}}{\{ab\}} + \frac{\partial\{xy\}}{\{xy\}}\right) \right] \quad (22)$$

Squaring and taking expectations,

$$V(r_G) = \frac{1}{16f\{ab\} \cdot \{xy\}} \left[4V\{ay\} + 4V\{xb\} + 8\,\mathrm{Cov}\,[\{ay\}, \{xb\}] \right.$$
$$+ (\{ay\} + \{xb\})^2 \left(\frac{V\{ab\}}{\{ab\}^2} + \frac{V\{xy\}}{\{xy\}^2} + 2\,\frac{\mathrm{Cov}\,[\{ab\}, \{xy\}]}{\{ab\} \cdot \{xy\}}\right)$$
$$- 4(\{ay\} + \{xb\})\left(\frac{\mathrm{Cov}\,[\{ay\}, \{ab\}] + \mathrm{Cov}\,[\{xb\}, \{ab\}]}{\{ab\}}\right.$$
$$\left.\left. + \frac{\mathrm{Cov}\,[\{ay\}, \{xy\}] + \mathrm{Cov}\,[\{xb\}, \{xy\}]}{\{xy\}}\right) \right] \quad (23)$$

It is now simplest to work in terms of covariances. From (9) we have the simple relations:

$$\mathrm{Cov}\,[\{ab\}, \{xy\}] = \frac{1}{f}\,[\{ax\} \cdot \{by\} + \{ay\} \cdot \{bx\}] \quad (24)$$

from which, by making different variates identical,

$$\left. \begin{aligned} \text{Cov}\left[\{ab\},\{ay\}\right] &= \frac{1}{f}\left[\sigma_a^2\{by\} + \{ab\}\cdot\{ay\}\right], \\[2mm] V\{ay\} &= \frac{1}{f}\left[\sigma_a^2\sigma_y^2 + \{ay\}^2\right], \quad \text{etc.} \end{aligned} \right\} \tag{25}$$

We next require the equations similar to (12) relating the variances and covariances of the four variates to h_1, h_2 etc. These are:

$$\left. \begin{aligned} \sigma_a^2 &= \frac{k}{2}\sigma_1^2, & \sigma_b^2 &= \frac{\sigma_1^2}{n}\left(1 + \frac{n-1}{2k}h_1^2\right), \\[2mm] \sigma_x^2 &= \frac{k}{2}\sigma_2^2, & \sigma_y^2 &= \frac{\sigma_2^2}{n}\left(1 + \frac{n-1}{2k}h_2^2\right) \\[2mm] \{ab\} &= \frac{1}{2}h_1^2\sigma_1^2, & \{xy\} &= \frac{1}{2}h_2^2\sigma_2^2 \\[2mm] \{ay\} &= \frac{1}{2}h_1h_2r_G\sigma_1\sigma_2 = \{xb\} \\[2mm] \{ax\} &= \frac{k}{2}r_P\sigma_1\sigma_2 \\[2mm] \{by\} &= \frac{1}{n}\left(r_P + \frac{n-1}{2k}h_1h_2r_G\right)\sigma_1\sigma_2 \end{aligned} \right\} \tag{26}$$

where as before { } indicates a covariance, σ_1^2 and σ_2^2 are the phenotypic variances of the two characters among individuals in an unselected population, and k takes the values 1 and 2, respectively, for full-sib families with both parents, and $\frac{1}{2}$-sib families with one parent measured.

Using (24), (25) and (26), we can express the covariances appearing in (23) in terms of h^2 etc., e.g.

$$\text{Cov}\left[\{ab\},\{xy\}\right] = \sigma_1^2\sigma_2^2\left[\frac{kr_P}{2n}\left(r_P + \frac{n-1}{2k}h_1h_2r_G\right) + \frac{1}{4}h_1^2h_2^2r_G^2\right]$$

$$\text{Cov}\left[\{ay\},\{ab\}\right] = \sigma_1^3\sigma_2\left[\frac{k}{2n}\left(r_P + \frac{n-1}{2k}h_1h_2r_G\right) + \frac{1}{4}h_1^3h_2r_G\right], \quad \text{etc.}$$

Making the relevant substitutions and simplifying, we obtain

$$V(r_G) = \frac{1}{f}\left[\frac{1}{2}(1-r_G^2)^2 + \frac{1}{2}(1-r_G^2)\left(\frac{1}{D} - \frac{r_Gr_P}{C}\right) + \frac{2k}{n}\left(\frac{r_G}{D} - \frac{r_P}{C}\right)^2\right.$$

$$\left. + \frac{1}{n}(1-r_G^2)\left\{\frac{k(1-r_P^2)}{C^2} - \frac{1}{2D} + \frac{r_Gr_P}{2C}\right)\right] \tag{27}$$

which is identical with (13),

In the same way it may be shown that the conditions of section 3 lead to a variance identical with equation (19), so that for any of the cases considered in this paper we may assume that the same variance applies, whether we calculate r_G by formula (1a) or (1b).

5. *The effect of variable family size.*

One is fortunate in a progeny test if all families yield the same number of progeny, and the question arises of dealing with variable n. In such a case, the amount of information supplied by the family mean is not usually proportional to the number of progeny, so that the commonly used weighting factor n is not the correct one. Kempthorne and Tandon (1953) have calculated the proper weights to be used in computing the regression or covariance of progeny mean on parent so that it has minimum sampling variance; but I found this paper rather difficult to follow in parts, and it may, therefore, be of interest to derive these weights by a simpler procedure, which will make their function clearer.

Suppose that we have families of n sibs, and that P, O and M are the mid-parent, individual progeny and progeny mean phenotypes of a single character, for any family. Let r_{OP} and r_{MP} be the correlations of an individual progeny and their family mean with mid-parent phenotype, and let r_{OO} be the intra-class correlation between sibs.

Then, for the variances of the regression coefficients of O and M on P, evidently

$$\left. \begin{array}{l} V(B_{OP}) = \sigma_O^2(1 - r_{OP}^2)/\sum (P - \bar{P})^2 \\ V(B_{MP}) = \sigma_M^2(1 - r_{MP}^2)/\sum (P - \bar{P})^2 \end{array} \right\} \qquad (28)$$

But $\sigma_M^2 = \sigma_O^2 [1 + (n - 1)r_{OO}]/n$, and Cov $(MP) =$ Cov (OP), so that $\sigma_M r_{MP} = \sigma_O r_{OP}$, and we can write:

$$V(B_{MP}) = \frac{\sigma_O^2(1 - r_{OO})}{\sum (P - \bar{P})^2} \left(\frac{1}{n} + \frac{r_{OO} - r_{OP}^2}{1 - r_{OO}} \right)$$

$$= \frac{\sigma_O^2(1 - r_{OO})}{w_n \sum (P - \bar{P})^2}, \qquad (29)$$

where $$w_n = \frac{n}{1 + nT},$$

$$\left. \begin{array}{l} \\ \end{array} \right\} \qquad (30)$$

and $$T = \frac{r_{OO} - r_{OP}^2}{1 - r_{OO}}$$

Only the denominator of (29) changes with family size n, so that the

weight of a regression coefficient based on families of size n is proportional to

$$W_n = w_n \sum (P - \overline{P})^2$$

To obtain the combined regression coefficient with minimum variance for variable n, we must first choose a common estimate of the parent mean $\overline{P}$, and then take a weighted average of the regression coefficients for each value of n, using the weight W_n. The appropriate estimate of $\overline{P}$ is obviously the weighted estimate

$$\tilde{P} = (\sum w_i P_i)/\sum w_i$$

summed over families, where P_i is the mid-parent value of family i, and w_i is the value of w_n, as defined in (30), for this family. Retaining subscripts n for families of n, and i for the i'th family, the average regression is

$$\overline{B} = \sum W_n b_n/\sum W_n = \sum w_i M_i(P_i - \tilde{P})/\sum w_i(P_i - \tilde{P})^2 \qquad (31)$$

These formulae apply equally if we are dealing with families of half-sibs correlated with their common parent P, and we then have the situation discussed by Kempthorne and Tandon (1953), in which $r_{OP} = \beta$ and $r_{OO} = \rho_1$ in their terminology. w_i is their weighting factor for families of n_i progeny, and (31) is identical with their weighted regression coefficient. We now see that this equation is simply the result of choosing a common parent mean $\tilde{P}$ and weighting the regression coefficients for each value of n in proportion to the reciprocals of their variances.

There are two cases to consider:
(1) families of $\frac{1}{2}$-sibs correlated with their common parent (as discussed by Kempthorne and Tandon)

$$r_{OP} = \tfrac{1}{2}h^2, \qquad r_{OO} = \tfrac{1}{4}h^2, \qquad \beta = \tfrac{1}{2}h^2,$$

(2) families of full sibs correlated with their mid-parent phenotype

$$r_{OP} = h^2/\sqrt{2}, \qquad r_{OO} = \tfrac{1}{2}h^2, \qquad \beta = h^2,$$

where β is in each case the regression of O on P. It follows that: for case (1)

$$T = \frac{h^2(1 - h^2)}{4 - h^2} = \frac{\beta(1 - 2\beta)}{2 - \beta} \qquad (32)$$

for case (2)

$$T = \frac{h^2(1 - h^2)}{2 - h^2} = \frac{\beta(1 - \beta)}{2 - \beta} \qquad (33)$$

These results are only strictly valid if we can ignore any non-additive genetic variance and any environmental correlations between progeny of the same family.

Kempthorne and Tandon point out that a value of T must be guessed in order to obtain the weighting factors w_i, but they give no indication how such a guess should be made (whether by gazing at a crystal or at the data), and this point seems to need clarification. If there are no better indications from other sources, one might suggest the use of formula (32) or (33) above, β being estimated from the unweighted regression of progeny on parent or mid-parent. In the example given by Kempthorne and Tandon, the unweighted estimate of β is 0.136, so that formula (20) gives $T = +0.053$, a value close to their guess of $T = 0.04$ (whose origin is not stated).

The effect of using any set of weights on the variance of the genetic correlation (13) will depend on its effect on the value of n turning up in formulae (12). Without proof, we state the following approximate results, which may be obtained by considering the effects of the various systems of weights on the R.H.S. of formulae (12).

Three sets of weights w_i may be considered, where the covariances in (1) are calculated as $\sum w_i M_i (P_i - \bar{P})$ in the terminology of the present section

(1) All $w_i = 1$, i.e. all family means given equal weight, regardless of the number of progeny in the family,

(2) $w_i = n_i$, i.e. progeny means weighted in proportion to number of progeny measured,

(3) $w_i = n_i/(1 + n_i T)$, for minimum variance of regression coefficients.

Then for n in formulae (13)–(15) and (19) we make the following substitutions:

$$
\left.
\begin{aligned}
&(1) \quad n = f \Big/ \sum \frac{1}{n_i} \\[2mm]
&(2) \quad n = \frac{1}{f-1}\left(\sum n_i - \frac{\sum n_i^2}{\sum n_i}\right) \\[2mm]
&(3) \quad n = \big[\big(\sum w_i\big)^2 - \sum (w_i^2)\big] \Big/ \left[\sum w_i \sum\left(\frac{w_i}{n_i}\right) - \sum\left(\frac{w_i^2}{n_i}\right)\right]
\end{aligned}
\right\} \quad (34)
$$

where $\sum$ indicates summation over families and f is the number of families. (3) reduces to (1) and (2), respectively in (34), when w_i is put equal to 1 and to n_i.

When using the weights $w_i = n_i/(1 + n_i T)$, the values of $T(T_1$ and $T_2)$ may be different for the two characters. If T_1 and T_2 do not

differ much, they can be averaged to give a single set of weights (w_i), without losing much information. If they differ enough to make appreciable differences in the weights for the two characters, then separate sets of weights should be used, and n in equation (3) of (34) may be taken, approximately, as the geometric mean of the values of n calculated separately from the weights of each character. If separate sets of weights are used, in estimating the genetic correlation by equation (1) the covariances should be calculated using the weight appropriate to the progeny variate in each case.

6. Examples.

In calculating the covariances from a progeny test, it may not be worth while to use Kempthorne and Tandon's weighting factors w_i , unless there is a good range of variation in progeny numbers. The genetic correlation may be estimated by either of formulae (1), and to estimate its variance we need only to apply formula (13) or the appropriate variation of it. For this purpose we need estimates of r_P , the phenotypic correction between the two characters in individuals of an unselected population, and of h_1^2 and h_2^2 , the fractions of the phenotypic variances of the two characters which can be attributed to additive genetic variations. These may be estimated from the regression coefficients in the progeny test, or from other sources, in the usual way. Then $C = h_1 h_2$ and $D = \frac{1}{2}(1/h_1^2 + 1/h_2^2)$.

As an example, tests on *Drosophila melanogaster* (Reeve and Robertson, 1953 and unpublished) show that in typical wild stocks wing and thorax length both have heritabilities of about 0.3, so that $C = D = 0.3$, and their phenotypic correlation is about 0.8, while the genetic correlation was estimated as 0.75.

Accepting these figures, in a progeny test with random mating, having f matings and n progeny per mating, formula (13) gives the variance of r_G as $0.4/f + 1.5/nf$. To obtain a standard error of ± 0.1 we need about 200 matings if $n = 1$, and about 60 matings if $n = 10$; and even if n is very large we still need about 40 matings to give the same accuracy.

In this case r_G is rather large, but as it declines the variance increases roughly in proportion to $(1 - r_G^2)$. Thus if both r_G and r_P are so small that they can be neglected in (13), and h_1^2 and h_2^2 are still both 0.3, the same progeny test gives $V(r_G) = (2.1/f) + (10/nf)$. This variance is over 6 times as great as before, so that we should now need 1200 matings with $n = 1$, and over 200 matings with n large, in order to obtain a standard error of ± 0.1.

7. *Summary.*

A formula is developed for the variance in large samples of the genetic correlation coefficient between two characters, estimated from the four parent-offspring covariances or correlations for the two characters. The variance is expressed in terms of the population values of the genetic and phenotypic correlations between the two characters and their heritabilities, and may be applied to progeny tests in which full-sib families and mid-parent values, or half-sib families and their common parent's phenotypes are measured. It can also be used for sex-limited characters or when covariances are calculated "within-sires". A modified formula is derived for the case of two progeny tests, one using selection and assortative mating for each character. The adjustments necessary when there are variable numbers of progeny per family are discussed, and an example illustrates the size of test necessary to estimate the genetic correlation with given accuracy, under various conditions.

It is shown that the variance is the same, whether the arithmetic mean or the geometric mean of the two covariances involving both characters is used in calculating the genetic correlation.

The limitations to be placed on the use of the variance formulae are discussed.

8. *Appendix*: selection and assortative mating.

On the assumption that we are dealing with strictly additive genetic effects, in a progeny test involving pair matings and full-sib progeny families, selection and assortative mating of parents for a given character both have the same effects on the progeny parameters, these effects depending simply on the factor $(1 + L)$ by which mid-parent variance is thereby multiplied (Reeve 1953).

Consider first the case of assortative mating of unselected parents, so as to introduce a phenotypic correlation L between mates for character 1. The path coefficient relationships between parents and offspring for two characters (subscripts 1 and 2), under these conditions, are shown in fig. 1. Primes indicate the parameters of the progeny generation, and otherwise the symbols have the notation used by Reeve (1953). It will be recalled that s is the additional correlation between the gametic genotypes of the two characters when these are carried in the same gamete as distinct from sister gametes. For unselected parents, $s = \frac{1}{2}r_G$, and ensures that the genetic correlation in zygotes remains constant from generation to generation, under random mating.

F_1 and F_2 are the correlations between uniting gametes for the two

characters, and are shown by dotted lines, since they are already implicit in the paths of the parents and must not be counted twice.

Fig. 1 gives immediately the following relationships:

$$\left. \begin{array}{l} F_1 = \tfrac{1}{2}Lh_1^2 \\ F_2 = \tfrac{1}{2}Lh_1^2 r_G^2 \end{array} \right\} \tag{35}$$

and for each character separately:

$$\left. \begin{array}{l} a^2 = 1/2(1 + F) \\ \sigma_{H'}^2 = (1 + F)\sigma_H^2 \\ \sigma_{P'}^2 = (1 + Fh^2)\sigma_P^2 \\ h'^2 = h^2(1 + F)/(1 + Fh^2) \\ a^2h'^2 = h^2/2(1 + Fh^2) \end{array} \right\} \tag{36}$$

Let Cov (P') and r_P' be the phenotypic covariance and correlation between characters 1 and 2 in a single progeny, and Cov $(12')$ and r_{12}'
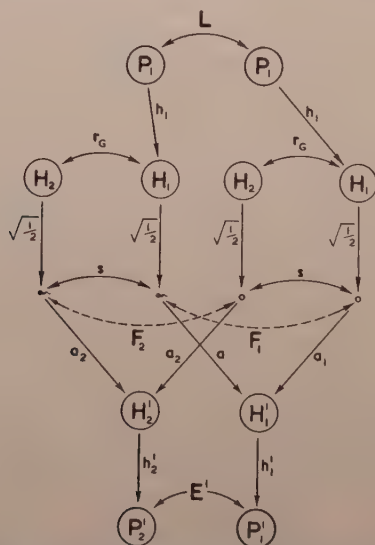
FIG. 1.   EFFECT OF ASSORTATIVE MATING FOR ONE CHARACTER ON RELATIONS OF PARENTS AND OFFSPRING FOR TWO CORRELATED CHARACTERS.

the covariance and correlation between characters 1 in one progeny and 2 in its sib of the same sex. Let E' and Cov (E') be the environmental components of r_P' and Cov (P'). We assume no environmental correlation between sibs. Then, from fig. 1:

$$r_P' = 2h_1'h_2'a_1a_2(s + \tfrac{1}{2}r_G + \tfrac{1}{2}r_G Lh_1^2) + E'$$
$$r_{12}' = 2h_1'h_2'a_1a_2(\tfrac{1}{2}r_G + \tfrac{1}{2}r_G Lh_1^2) \tag{37}$$

In terms of covariances, after substituting from (36), these become:

$$\text{Cov}\,(P') = (r_P + \tfrac{1}{2}h_1 h_2 r_G Lh_1^2)\sigma_1\sigma_2$$
$$\text{Cov}\,(12') = \tfrac{1}{2}h_1 h_2 r_G(1 + Lh_1^2)\sigma_1\sigma_2 \tag{38}$$

where all the parameters on the R.H.S. refer to the random-mating population. In obtaining the first equation of (38), we note that non-random mating of parents does not affect Cov (E'), so that

$$\text{Cov}\,(E') = (r_P - h_1 h_2 r_G)\sigma_1\sigma_2$$

For the covariance $\{by\}$ between the means of n sib progeny of the two characters, we obtain:

$$\{by\} = \frac{1}{n}\,[\text{Cov}\,(P') + (n-1)\,\text{Cov}\,(12')]$$

$$= \frac{1}{n}\,[r_P + \tfrac{1}{2}(n-1)h_1 h_2 r_G + \tfrac{1}{2}Lh_1^3 h_2 r_G]\sigma_1\sigma_2 \tag{39}$$

By an easy extension of fig. 1 it may be seen that, for two sibs and a single character we have

$$\sigma_{P'}^2 = (1 + Fh^2)\sigma_P^2$$
$$\text{Cov}\,(P'P') = 2h''^2 a^2(\tfrac{1}{2} + F)\sigma_{P'}^2 \tag{40}$$

For the family mean of n sib progeny $(\overline{P}')$ these equations yield

$$\sigma_{\overline{P}'}^2 = \frac{1}{n}\,[1 + \tfrac{1}{2}(n-1)h^2 + nFh^2]\sigma_P^2 \tag{41}$$

On interchanging subscripts 1 and 2 in the above formulae we obtain the corresponding results for assortative mating of character 2.

Since assortative mating and selection of parents have the same effect on the progeny parameters, equations (39) and (41) apply to either system or to the combined effects of both, provided that their total effect is to multiply mid-parent variance for the character selected by $(1 + L)$. Thus, on putting in the appropriate subscripts, the above equations give the progeny variance and covariance formulae needed for deriving equations (18).

As a corollary, we may deduce that selection of parents so as to

multiply the mid-parent variance of character 1 by $(1 + K)$ converts the parameter s of fig. 1 into:

$$s_0 = \tfrac{1}{2}r_G/[(1 + \tfrac{1}{2}Kh_1^2)(1 + \tfrac{1}{2}Kh_1^2r_G^2)]^{1/2} \qquad (42)$$

It should be emphasised that the equations referring to non-random mating are only strictly valid when all the genetic variance is additive and there are no environmental correlations between parent and off-spring (the first of these conditions is, no doubt, rarely satisfied in practice). The amount of disturbance to be expected with any departure from these conditions is quite unknown.

BIBLIOGRAPHY

Hazel, L. N.   The genetic basis for constructing selection indices.   *Genetics 28,* 476–90 (1943).

Kempthorne, O. and O. B. Tandon.   The estimation of heritability by regression of offspring on parent.   *Biometrics 9,* 90–100 (1953).

Morley, F. H. W.   Selection for economic characters in Australian Merino sheep. (1) Estimates of phenotypic and genetic parameters.   N. S. W. Dep. Agric. Sci. Bull. No. 73 (1951).

Reeve, E. C. R.   Studies in quantitative inheritance.   III Heritability and genetic correlation in progeny tests using different mating systems.   *J. Genet. 51,* 520–42 (1953).

Reeve, E. C. R. and F. W. Robertson.   Studies in quantitative inheritance. II Analysis of a strain of *Drosophila melanogaster* selected for long wings.   *J. Genet. 51* 276–316 (1953).

TESTS FOR LINEAR TRENDS IN PROPORTIONS AND FREQUENCIES

P. Armitage

Statistical Research Unit of the Medical Research Council,
London School of Hygiene and Tropical Medicine

1. *Introduction*

One frequently encounters data consisting of a series of proportions, occurring in groups which fall into some natural order. The question usually asked is then not so much whether the proportions differ significantly, but whether they show a significant trend, upwards or downwards, with the ordering of the groups. In the data shown in Table 1, for instance, the usual test for a 2×3 contingency table yields a χ^2 equal to 7.89 on 2 degrees of freedom, corresponding to a probability of about 0.02. But this calculation takes no account of the fact that the carrier rate increases with the tonsil size, and it is reasonable to believe that a test specifically designed to detect a trend in the carrier rate as the tonsil size increases would show a much higher degree of significance.

TABLE 1

Relationship between nasal carrier rate for Streptococcus pyogenes and size of tonsils, among 1398 children aged 0–15 years. (Data from Drs. M. C. Holmes and R. E. O. Williams, summarised by Holmes and Williams, 1954)

	Present, but not enlarged +	Enlarged tonsils		Total
		++	+++	
Carriers	19	29	24	72
Non-carriers	497	560	269	1326
	516	589	293	1398
Carrier-rate	0.0368	0.0492	0.0819	

No originality is claimed for the tests discussed in this paper. They will be familiar to many statisticians, and may be derived as particular cases of procedures already published for contingency tables with any

number of rows and columns. Since the situation in which one of the classifications in a contingency table is a dichotomy (so that the data form a series of proportions) occurs so frequently, it is hoped that an explicit discussion of this case may be of interest.

We shall regard the data as forming a $2 \times k$ contingency table, and use the following notation:

Column

	1	2	3	$\cdots$	k	Total
Row 1	n_1	n_2	n_3	$\cdots$	n_k	t
Row 2	$N_1 - n_1$	$N_2 - n_2$	$N_3 - n_3$	$\cdots$	$N_k - n_k$	$T - t$
	N_1	N_2	N_3	$\cdots$	N_k	T

The proportion of individuals in the i-th column, which fall into the first row, is denoted by $p_i = n_i/N_i$, and the overall proportion is $P = t/T$. In summations (which are always over the k columns), we shall omit the suffix i. Thus, $\sum Nx$ will denote $\sum_{i=1}^{k} N_i x_i$.

2. A test based on scores

To measure and test the significance of the trend in the p_i, a natural procedure is to allot a score x_i to the i-th column ($x_1 < x_2 < \cdots < x_k$), and to perform some sort of regression analysis of p on x. In addition to the column scores x_i, let us allot to each of the T observations a row score, y, taking the values $y = 1$ for each observation in Row 1, and $y = 0$ for Row 2. Then the mean value of y for the i-th column is clearly $n_i/N_i = p_i$, and the overall mean of y is $t/T = P$. Thus, a regression analysis of y on x will be equivalent to one of p on x (p_i being weighted in proportion to N_i). The T values of y could now be subjected to a formal analysis of variance, between and within columns, as follows:

	Degrees of freedom	Sum of squares
Between columns		
Due to linear regression	1	S_1
Departures from linearity	$k - 2$	S_2
	$k - 1$	$S_1 + S_2$
Within columns	$T - k$	S_3
Total	$T - 1$	$S_1 + S_2 + S_3$

where
$$S_1 = \{\sum Np(x - \bar{x})\}^2 / \sum N(x - \bar{x})^2, \qquad (1)$$

$$S_1 + S_2 = \sum N(p - P)^2,$$

$$S_3 = \sum Np(1 - p),$$

$$S_1 + S_2 + S_3 = TP(1 - P), \qquad (2)$$

and
$$\bar{x} = \sum Nx/T.$$

Consider first the problem of testing for general heterogeneity between columns. As in the usual model for the analysis of variance, we assume that in repeated sampling the column totals N_i are fixed. The null hypothesis is that the expected value of y (and hence of p_i) is the same for all columns. The usual analysis of variance test is to calculate the variance ratio $\{(S_1 + S_2)/(k - 1)\}/\{S_3/(T - k)\}$. However, with a variate such as y, taking only the values 0 or 1, the normal theory is strictly valid only for large samples, and in these circumstances a number of alternative approximate tests are available. In particular the usual formula for χ^2 on $k - 1$ degrees of freedom can be expressed as

$$(S_1 + S_2)/\{(S_1 + S_2 + S_3)/T\}. \qquad (3)$$

Here the denominator is taken from the "Total" row in the analysis of variance table, but with the divisor T instead of the total degrees of freedom $T - 1$. In all these alternative tests, the tabulated χ^2 distribution is strictly valid only asymptotically for large sample sizes, and the tests become equivalent as the N_i increase, provided that the null hypothesis is true.

Similarly, to test the significance of the regression, the usual analysis of variance procedure would be to compare S_1 with S_3 (or perhaps with $S_2 + S_3$ if we ignored the possibility of departures from linearity). An alternative test, equivalent in large samples if the null hypothesis is true, is to calculate

$$\chi_0^2 = S_1/\{(S_1 + S_2 + S_3)/T\}, \qquad (4)$$

which is distributed approximately as χ^2 on 1 degree of freedom. If we wish to calculate confidence limits for the regression coefficient, assuming that the true value might differ from zero, we should use $S_3/(T - k)$ as an estimate of variance rather than $(S_1 + S_2 + S_3)/T$.

Which of the various alternative criteria follows most closely its assumed sampling distribution, for small samples, is a matter for further study; (see the Appendix, §6). In the meantime, there seems little objection to the use of (4). This criterion is equivalent to that

proposed by Yates (1948) for contingency tables with any number of rows and columns. For $k = 2$, it is equivalent to the usual χ^2 criterion for 2×2 tables (without continuity correction). For $k > 2$, a comparison of (3) and (4) shows that χ_0^2 is a part of the total χ^2, the difference between the two values representing departures from linearity, and having $k - 2$ d.f.

Denoting by b the estimated regression coefficient of y on x, and by $V(b)$ the estimated* sampling variance of b on the null hypothesis, we find that

$$b = \frac{\sum Np(x - \bar{x})}{\sum N(x - \bar{x})^2} = \frac{T \sum nx - t \sum Nx}{T \sum Nx^2 - (\sum Nx)^2} , \qquad (5)$$

$$V(b) = \frac{P(1 - P)}{\sum N(x - \bar{x})^2} = \frac{t(T - t)}{T\{T \sum Nx^2 - (\sum Nx)^2\}} , \qquad (6)$$

and, from (1), (2) and (4),

$$\chi_0^2 = \frac{b^2}{V(b)} = \frac{T\{T \sum nx - t \sum Nx\}^2}{t(T - t)\{T \sum Nx^2 - (\sum Nx)^2\}} \qquad (7)$$

on 1 degree of freedom.

The calculations cannot be performed until the scores x_i have been chosen. In the absence of any *a priori* knowledge of the type of trend to be expected, it seems reasonable to choose the x_i to be equally-spaced, and it will often be convenient to have them centred around zero. This is the procedure advocated by Yates. Thus, for k columns, we should choose $x_1 = -\frac{1}{2}(k - 1)$, $x_2 = -\frac{1}{2}(k - 3)$, $\cdots$, $x_k = \frac{1}{2}(k - 1)$. The choice of scores is discussed further in a later section. It should be emphasized that, whatever scoring system is chosen, the validity of the significance test is not affected; that is, if the null hypothesis is true, a value of χ_0^2 significant at the $\alpha\%$ level will occur only about α times out of 100.

As an example, using the data of Table 1, we shall allot equally-spaced scores as follows: $x_1 = -1$, $x_2 = 0$, $x_3 = 1$. We obtain

$$\sum nx = 5, \qquad \sum Nx = -223, \qquad \sum Nx^2 = 809,$$

whence, from (5), (6) and (7),

$$b = 0.02131,$$

$$V(b) = 0.000063160; \qquad \sqrt{V(b)} = 0.00795,$$

and $\qquad\qquad \chi_0^2 = 7.19$ on 1 d.f. $\qquad (P = 0.007).$

*In repeated sampling with both sets of marginal totals fixed, the expression (6) is $(T - 1)/T$ times the *exact* variance of b. This can be shown from results given by Haldane (1940).

The test for trend indicates, as expected, a considerably higher degree of significance than the total χ^2 of 7.89 on 2 d.f. The test for departures from linear regression gives $\chi^2 = 7.89 - 7.19 = 0.70$ on 1 d.f., which is non-significant. In this particular example, the association between carrier rate and tonsil size may be due to the association of both factors with the age or social class of the child.

Yates (1948) points out that the same formula for χ_0^2 is obtained whether one considers the regression of row score on column score, or that of column score on row score. Now, when there are only two rows, a test for the regression of column score on row score is equivalent to a test for the difference between the mean column score for the first row and that for the second row. For some types of data, particularly where the row totals are fixed beforehand, it will be more natural to think of the χ_0^2 test in this way, rather than in terms of the regression of p on x. In the data shown in Table 2, for instance, the row totals, 32 and 32, were fixed by the experimental design, and it seems more natural to ask whether the mean scores in the two treatment groups differ significantly, rather than whether the proportion of patients in group A, in each column, shows a linear trend with the score. In this example, the total $\chi^2 = 5.91$ on 3 d.f. ($P = 0.12$), whereas $\chi_0^2 = 5.26$ on 1 d.f. ($P = 0.02$), showing a fairly definite improvement in group A as compared with group B.

TABLE 2

Changes in size of ulcer crater, 3 months after start of treatment, for patients in two treatment groups (From Table IV of Doll and Pygott, 1952)

Treatment group	Number of cases with crater				Total
	Larger	Less than 2/3 healed	2/3 or more healed	Healed	
A	6	4	10	12	32
B	11	8	8	5	32
	—	—	—	—	—
	17	12	18	17	64
Score, x_i	−1.5	−0.5	+0.5	+1.5	

A test criterion exactly equivalent to χ_0^2 has been used in genetical applications by Fisher and Ford (1947, p. 163) and by Holt (1948, p. 148). A recent example of the use of this test, in a 2 × 3 table, is given by Grüneberg (1955). He compares the proportions of animals in two

stocks which show some effect on 0, 1 or 2 sides of the body. The formula for χ^2 given by C. A. B. Smith in the Appendix to Grüneberg's paper is equivalent to our (7). The more general problem in which more than two stocks are compared could be treated by Yates's methods.

3. Trends in frequencies

If $P = t/T$ is very small, we may substitute $T/(T - t) \sim 1$ in (7). Defining $e_i = tN_i/T$, the "expected" frequency corresponding to the observed frequency n_i, we find from (7) that

$$\chi_0^2 = \frac{\{\sum x(n - e)\}^2}{\sum ex^2 - (\sum ex)^2/t}. \tag{8}$$

The numerator of (8) is the square of the cross-product, U, of the scores x_i with the discrepancies $n_i - e_i$. The denominator is equal to $\sum e (x - \bar{x})^2$, where $\bar{x} = \sum ex/t$, i.e. a weighted sum of squares of the x_i about their mean, the weights being the expected numbers. The test is thus based entirely on the frequencies in the first row, and is clearly valid only when the sampling errors of the frequencies in the second row are relatively negligible. The frequencies in the first row may be thought of as those occurring in a sample of size t from a multi-nomial distribution. The denominator of (8) is then obtained directly as the variance of U in repeated sampling, with t and the expected frequencies e_i kept constant. The expression (8) may thus be written as $\chi_0^2 = U^2/V(U)$.

In Table 3, the expected frequencies e_i, have been obtained by sub-dividing the total number of maternal deaths, 127, in proportion to the number of mothers at risk during each of the eight periods. The last line of Table 3 suggests, perhaps, a slight tendency for the maternal mortality rates to fall. The scores, x_i, have been taken as the mid-points of the different periods, minus 1900. The total χ^2, calculated from the observed and expected frequencies is 3.91 on 7 d.f. ($P = 0.79$); even if the whole of this quantity were ascribed to regression it would barely reach the 5% level of significance on 1.d.f. In fact, application of (8) gives $\chi_0^2 = 1.27$ on 1 d.f. ($P = 0.26$). The data, therefore, do not provide any evidence for a gradual decline in maternal mortality amongst women of this particular parity and age-group.

4. Kendall's rank correlation test

An alternative approach to data like those in Tables 1 and 2 is to apply rank correlation methods (Kendall, 1948; Stuart, 1953). In Table 1, for instance, we could regard the 1398 children as being ranked

TABLE 3
Maternal mortality in New South Wales, for primiparae aged 40 and over. (From Tables I, II and III of Wilcocks and Lancaster, 1951)

	1894–1900	1901–1907	1908–1910	1911–1920
x_i	−2.5	4.5	9.5	16.0
Number of mothers, N_i	346	454	272	1133
Deaths				
Observed, n_i	10	9	3	23
Expected, e_i	6.603	8.664	5.191	21.621
Maternal mortality rate,				
per 1,000	28.90	19.82	11.03	20.30

	1921–1930	1931–1937	1938–1942	1943–1948	Total
x_i	26.0	34.5	40.5	46.0	
Number of mothers, N_i	1546	909	699	1296	6655
Deaths					
Observed, n_i	32	17	13	20	127
Expected, e_i	29.503	17.347	13.339	24.732	127.000
Maternal mortality rate,					
per 1,000	20.70	18.70	18.60	15.43	

in two ways. In the first ranking (corresponding to the rows of Table 1), 1326 individuals are tied with a rank of $(1 + 1326)/2 = 663.5$, and the remaining 72 are tied with a rank of $1326 + (1 + 72)/2 = 1362.5$. In the second ranking (for columns), 516 are tied with a rank of $(1 + 516)/2 = 258.5$, 589 are tied with a rank of $516 + (1 + 589)/2 = 811.0$, and 293 are tied with a rank of $516 + 589 + (1 + 293)/2 = 1252.0$. To test for a tendency for the carrier-rate to increase or decrease with tonsil size, we could apply the usual techniques of rank correlation, making allowance for the considerable number of ties.

To calculate Kendall's statistic, S (§1.9 of his book), we form the sum of products of each frequency in the second row with the frequencies above and to the right of it, and subtract the sum of products of each frequency in the first row with those below and to the right of it. Thus, in the notation previously used:

$$S = (N_1 - n_1)(n_2 + n_3 + \cdots + n_k) + (N_2 - n_2)(n_3 + \cdots + n_k)$$

$$+ \cdots + (N_{k-1} - n_{k-1})n_k - n_1\{(N_2 - n_2) + \cdots + (N_k - n_k)\}$$

$$- \cdots - n_{k-1}(N_k - n_k). \tag{9}$$

When the null hypothesis is true (i.e. there is no association), the variance of S is (writing Kendall's (4.5) in the present notation),

$$V(S) = \frac{t(T - t)}{3T(T - 1)} (T^3 - \sum N^3). \tag{10}$$

(Stuart (1953) considers inequalities for the variance when the null hypothesis is not true.) A test for association is thus provided by

$$\chi_1^2 = S^2/V(S) \quad \text{on} \quad 1 \text{ d.f.} \tag{11}$$

If $k = 2$, χ_1^2 is equal to $(T - 1)/T$ times the usual χ^2 for a 2×2 table (without continuity correction). This factor is of no great importance, in view of the asymptotic nature of the assumed χ^2 distribution.

At first sight the approach of §2 seems to bear little relationship to that of the present section. In point of fact the two methods are quite closely related. It is known (Hemelrijk, 1952) that when one of the classifications in a rank correlation table is a dichotomy, Kendall's test based on S is equivalent to Wilcoxon's test for the sum of the ranks in one of the sub-groups (see Kruskal and Wallis, 1952, for references). This, in turn, is equivalent to a test for the difference between the mean ranks in the two sub-groups, since the overall sum of ranks is constant. This difference would be the same as the difference in mean column scores, discussed in §1, if we chose the score for each column to be equal to the mid-rank for that column. Thus, we should have $x_1 = (1 + N_1)/2$, $x_2 = (1 + 2N_1 + N_2)/2$, $x_3 = (1 + 2N_1 + 2N_2 + N_3)/2$, etc. It would, therefore, not be surprising if the χ_1^2 test were closely related to the χ_0^2 test with the x_i chosen in this way, or at least chosen so as to be linearly related to these values. It is not difficult to show directly that this is so.

Rearranging the terms in (9), and writing $p_i = n_i/N_i$, we find that

$$S = \sum nx, \tag{12}$$

where $x_i = N_1 + N_2 + \cdots + N_{i-1} - N_{i+1} - \cdots - N_k$

$$= (1 + 2N_1 + \cdots + 2N_{i-1} + N_i) - (T + 1). \tag{13}$$

These scores x_i are linearly related to the mid-ranks given above, and it can easily be verified that

$$\sum Nx = 0 \quad \text{and} \quad \sum Nx^2 = (T^3 - \sum N^3)/3. \tag{14}$$

Hence, from (7) and (14),

$$\chi_0^2 = \frac{3T^2 S^2}{t(T - t)(T^3 - \sum N^3)}$$

$$= \{T/(T - 1)\}\chi_1^2, \quad \text{from (10) and (11)} \tag{15}$$

When the N_i are equal, the x_i are clearly equally-spaced. The rank correlation test is then equivalent to the regression test with equally-spaced scores, except for the factor $T/(T-1)$ in (15). As already stated, this factor is asymptotically unimportant. The tests would have been exactly equivalent if, in the formula (4) from which (7) is derived, the total degrees of freedom, $T-1$, had been used as a divisor in the denominator, instead of T. As we have seen, when $k=2$, χ_0^2 agrees with the usual χ^2 for a 2×2 table, whereas χ_1^2 differs from it by a factor $(T-1)/T$.

As examples of the rank correlation test, formulae (9)–(11) have been applied to the data shown in Tables 1 and 2. For Table 1,

$$S = 16229, \qquad V(S) = 38{,}543{,}560.2,$$

and
$$\chi_1^2 = 6.83 \qquad (P = 0.009),$$

as compared with $\chi_0^2 = 7.19$. For Table 2,

$$S = 330 \qquad V(S) = 20720.25$$

$$\chi_1^2 = 5.26 \qquad (P = 0.02),$$

as compared with $\chi_0^2 = 5.26$ (the exact agreement being coincidental).

5. Choice of test

Since the rank correlation test has been shown to be equivalent (apart from the factor $T/(T-1)$) to the regression test, with a particular choice of scores depending on the N_i, the decision whether to use χ_1^2 or χ_0^2 reduces to a choice of the most suitable system of scoring. In most situations there will be no prior reason to expect any particular type of relationship, and it is difficult to formulate any general advice.

If the columns are defined by a measurement, like age, it will often be reasonable to choose scores linearly related to the values assumed by the measurement, taking mid-points of groups where necessary (as in Table 3).

If the columns are defined by a qualitative classification as in Table 1, the choice is more arbitrary. If the problem is primarily thought of as a trend in proportions in well-defined ordered groups, the regression method with equally-spaced x_i seems the most appropriate. An estimate is obtained of the mean change in p_i from group to group, and one avoids the use of scores depending on the N_i which may be difficult to interpret. If, on the other hand, the grouping by columns is arbitrary, there may be little virtue in using equally-spaced x_i, and the rank correlation method is perhaps the more objective. Fortunately, the two tests will usually give fairly close results.

It may be of interest to conclude with a historical note. The reader will find a number of sets of data suitable for analysis by the methods outlined here, in two papers by Karl Pearson (1909, 1910). In the first paper, Pearson considered situations in which the columns corresponded to a numerical variate; the rows were assumed to represent a dichotomy of an underlying normal variate and the method provided an estimate of the hypothetical correlation coefficient (sometimes referred to as "biserial r"). In the second paper, the method was extended for data in which the columns were qualitatively defined, but might still be ordered; an estimate of the hypothetical correlation ratio ("biserial η"), not dependent on the ordering, was obtained, and the trend was assessed by inspection. These methods have largely fallen into disuse, partly because of difficulties in determining the sampling errors of the coefficients, and partly because the existence of a normal variate underlying the dichotomy by rows was not generally accepted.

6. *Appendix. Exact distribution of* χ_0^2 *on the null hypothesis*

The exact distribution of χ_0^2 has been determined, by enumeration of all possible results, for the case where $k = 3$, $N_1 = N_2 = N_3 = 10$, and the n_i each follow the binomial distribution $(\frac{1}{2} + \frac{1}{2})^{10}$. The probabilities with which χ_0^2 exceeds various tabulated percentiles of the χ^2 distribution on 1 d.f. are shown in Table 4. This table also shows the cumulative distribution of an alternative test criterion,

$$\chi_2^2 = S_1 / \{(S_2 + S_3)/(T - k)\},$$

in the notation of §2. The formula for χ_2^2 differs from that for χ_0^2, (4), in having as denominator the mean square about regression. The two test criteria are connected by the relationship

$$\chi_2^2 = (T - 2)\chi_0^2/(T - \chi_0^2),$$

$$= 28\chi_0^2/(30 - \chi_0^2)$$

since $T = 30$.

Although the expected frequencies in this example are as low as 5, Table 4 shows (a) that there is little to choose between the two tests up to about the 5% level of significance, and (b) that the distribution of either test criterion agrees well with the theoretical χ^2 distribution between the 50% and 5% points. The appreciable discrepancy at the lower end of each distribution is due to there being a probability of 0.176 that $\chi_0^2 = \chi_2^2 = 0$. It would be dangerous to generalize from this example alone, but the results are at least encouraging.

Values of χ^2	Cumulative probability		
	Tabulated	χ_0^2	χ_2^2
0–	1.000	1.000	1.000
0.0^3157-	0.990	0.824	0.824
0.0^3628-	0.980	0.824	0.824
0.0^2293-	0.950	0.824	0.824
0.0158–	0.900	0.824	0.824
0.0642–	0.800	0.824	0.824
0.148–	0.700	0.824	0.824
0.455–	0.500	0.504	0.504
1.074–	0.300	0.264	0.264
1.642–	0.200	0.263	0.263
2.706–	0.100	0.116	0.116
3.841–	0.050	0.042	0.044
5.412–	0.020	0.014	0.042
6.635–	0.010	0.012	0.013
10.827–	0.0010	0.0005	0.0028

I am indebted to Professor A. Bradford Hill and Dr. J. O. Irwin for commenting on the first draft of this paper; to Dr. M. C. Holmes and Dr. R. E. O. Williams for permission to quote, in Table 1, details not appearing in their paper; and to Miss Irene Allen for computing assistance.

Since this paper was accepted for publication, the regression test based on χ_0^2 has been discussed by W. G. Cochran (1954), *Biometrics* *10*: 417–451 §§6.2, 6.3.

REFERENCES

Doll, R. and Pygott, F. (1952). Factors influencing the rate of healing of gastric ulcers. *Lancet*, i, 171–182.

Fisher, R. A. and Ford, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity*, *1*, 143–174.

Grüneberg, H. (1955). Genetical studies on the skeleton of the mouse. XV. The interaction between major and minor variants. (With an Appendix by C. A. B. Smith). *J. Genet.* (in press)

Haldane, J. B. S. (1940). The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small. *Biometrika*, *31*, 346–355.

Hemelrijk, J. (1952). Note on Wilcoxon's two-sample test when ties are present. *Ann. Math. Statist.*, *23*, 133–135.

Holmes, M. C. and Williams, R. E. O. (1954). The distribution of carriers of *Streptococcus pyogenes* among 2,413 healthy children. *J. Hyg., Camb.*, *52*, 165–179.

Holt, S. B. (1948). The effect of maternal age on the manifestation of a polydactyl gene in mice. *Ann. Eugen., Lond., 14,* 144–157.

Kendall, M. G. (1948). *Rank correlation methods.* London: Griffin.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion analysis of variance. *J. Amer. Statist. Ass., 47,* 583–621.

Pearson, K. (1909). On a new method of determining correlation between a measured character *A*, and a character *B*, of which only the percentage of cases wherein *B* exceeds (or falls short of) a given intensity is recorded for each grade of *A*. *Biometrika, 7,* 96–105.

Pearson, K. (1910). On a new method of determining correlation, when one variable is given by an alternative and the other by multiple categories. *Biometrika, 7,* 248–257.

Stuart, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika, 40,* 105–110.

Wilcocks, W. J. and Lancaster, H. O. (1951). Maternal mortality in New South Wales with special reference to age and parity. *J. Obstet. Gynaec., Brit. Emp., 58,* 945–960.

Yates, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. *Biometrika, 35,* 176–181.

AN EXAMPLE OF THE TRUNCATED POISSON DISTRIBUTION

D. J. Finney* and G. C. Varley

University of Oxford

1. *The problem and the data*

The female knapweed gall-fly, *Urophora jaceana*, lays its eggs in batches within unopened flower-heads of black knapweed, *Centaurea nemoralis*. The eggs are easily seen and counted when the flower-head is split open. The second instar larva hatches from the egg, moves a short distance within the flower-head, and produces a hard gall-cell in which occur all further stages of development up to emergence of the adult fly.

As part of his intensive study of population balance in the gall-fly, Varley (1947, pp. 158–161) wished to estimate the total mortality that occurs between oviposition and gall formation. He had records from samples of flower-heads in 1935 and 1936, first for numbers of eggs per flower-head and at a later date for numbers of gall-cells per flower-head. The sampling procedure is not under discussion here. Table 1 contains the results, each sample relating to different flower-heads since the process of counting is destructive; the number of 'empty' flower-heads is omitted because a multitude of causes not relevant to the investigation may secure that no eggs are laid. Some eggs or larvae will fail to produce gall-cells, but each gall-cell observed corresponds to only one egg.

These data will be used here as examples of the utility of the truncated Poisson distribution, which has recently been the subject of several papers. Analysis with the aid of that distribution leads to some modification of Varley's previous conclusions from the data, at least in respect of the strength of evidence for the occurrence of competition within the flower-head.

2. *The earlier analysis*

Varley inquired whether the mortality between oviposition and gall formation was independent of the number of eggs per flower-head, his expectation being that relatively more deaths would occur in the heavily populated flower-heads. Suppose that failure of eggs to produce gall-cells occurs entirely randomly in a proportion θ of eggs. Then, of flower-heads that initially have x eggs, the proportion that later

---

*Now at the University of Aberdeen.

TABLE 1

Observed frequencies of eggs and gall-cells

No. of eggs or gall-cells	Frequency of flower-heads			
	1935		1936	
	Eggs	Gall-cells	Eggs	Gall-cells
1	29	287	22	90
2	38	272	18	96
3	36	196	18	57
4	23	79	11	26
5	8	29	9	10
6	5	20	6	4
7	5	2	3	5
8	2	0	0	0
9	1	1	1	1
10	0	0	0	0
11	0	0	0	0
12	1	0	0	0
>12	0	0	0	0
Total	148	886	88	289

contain y gall-cells will be determined by the binomial distribution as

$$\binom{x}{y}\theta^{x-y}(1 - \theta)^{y} \qquad (y \leq x).$$

Using the observed egg distribution from Table 1, Varley calculated the gall-cell distribution for various trial values of θ and compared these with the observed distribution in the same year by means of χ^2. He estimated θ as that value which minimized χ^2, and obtained a variance for the estimate from the rate of change of χ^2 in the neighbourhood of the minimum. His estimates were 0.289 ± 0.022 for 1935 and 0.323 ± 0.035 for 1936. In 1935, his minimum χ^2 was sufficiently large to suggest that the mortality was not operating at random but that possibly it was higher for the flower-heads with many eggs than for those with few.

The minimization of χ^2 is well-known to be a fully efficient procedure for estimating a parameter from large samples. In the present problem, however, it fails to take account of the fact that the distribution of eggs used in calculating the expected frequencies for the cell distribution is itself based upon a sample. If these expected frequencies were expressed as functions solely involving unknown parameters, or if the

observations on gall-cells had been made at a later date on the same flower-heads as were used for the egg records,* the method would be appropriate. In the particular circumstances of Varley's data, though the minimization of χ^2 is still a valid method of estimating the mortality rate, the values of χ^2 will be increased by neglect of the sampling variation of the egg distribution. Consequently, evidence for heterogeneity of mortality will be exaggerated and the standard error of the estimate of θ, obtained from consideration of the rate of change of χ^2, will be biased downwards. If the number of flower-heads used for the egg records were much larger than that for gall-cells, sampling errors in the egg distribution could safely be ignored; for both years, however, the egg sample was substantially smaller than the gall-cell sample. A process complementary to that used by Varley, namely taking the gall-cell distribution as fixed by the sample and calculating χ^2 by comparison of the egg distribution with a theoretical one giving the right gall-cell distribution ought to give more trustworthy results when the latter is based on so much the larger sample, but this would introduce new difficulties because some flower-heads containing eggs will yield no gall-cells and because no egg distribution could be found to give exactly a specified gall-cell sample as expected frequencies. A more satisfactory method is to treat the two samples similarly, first estimating comparable mean numbers of eggs and gall-cells and then estimating the survival rate, $(1 - \theta)$, as the ratio of these.

3. *The truncated Poisson*

The mean numbers of eggs and gall-cells per flower-head taken directly from Table 1 are not comparable, on account of the omission of zeros: some flower-heads containing eggs will produce no gall-cells, so that the ratio of these means would tend to overestimate the survival rate. Further progress seems to demand the use of a reasonably simple parametric model of the situation. The frequency distribution of x, the number of eggs per flower-head, in the two samples reported in Table 1, has the appearance of a Poisson distribution truncated by the absence of observations for $x = 0$, and it is of interest to see whether the data can be adequately described in this way. The most obvious suggestion, that a Poisson distribution is generated by the random deposition of single eggs laid on such flowers as are at the right stage of development, is untenable here since it is known that a female normally lays several eggs at a time and that one flower-head rarely receives eggs from more than one female; the distribution observed must therefore

---

*In this investigation, it was not possible to count eggs without destroying the flower-head, but in analogous studies of other organisms such an observational programme might be adopted.

approximate closely to that of the size of egg batches. Nevertheless, the Poisson model will be shown to operate very satisfactorily.

If the distribution of x is Poisson with mean λ and if mortality occurs randomly and independently of x, then the distribution of y, the number of gall-cells per flower-head, is easily shown to be also Poisson, the mean being

$$\mu = \lambda(1 - \theta). \tag{1}$$

Now for a Poisson distribution with the first term omitted,

$$P(x) = \frac{\lambda^x}{x!(e^\lambda - 1)}. \tag{2}$$

The maximum likelihood estimate of λ is obtained (David & Johnson, 1952) by equating the mean value of x to its expectation, and so is $\hat\lambda$, the solution of

$$\bar x = \lambda/(1 - e^{-\lambda}). \tag{3}$$

Rider (1953) independently obtained the same equation and provided a brief table to help in evaluating $\hat\lambda$. Moreover, by the usual maximum likelihood procedure, the variance of $\hat\lambda$ is found to be asymptotically of the form

$$V(\hat\lambda) = \frac{\lambda^2}{N\bar x(\lambda + 1 - \bar x)} \tag{4}$$

where N is the number of observations on x from which $\bar x$ is formed. Cohen (1954) has obtained more general formulae, of which (3) and (4) are special cases. Equations similar to (2), (3), (4), with y, μ in place of x, λ are also required.

David and Johnson remarked on the impossibility of obtaining an explicit expression for $\hat\lambda$ as a function of $\bar x$, with the implication that this is a serious disadvantage of the method of estimation. In practice, equation (3) can be solved rapidly by iterative or interpolatory processes and a table for direct reading of $\hat\lambda$ and $NV(\hat\lambda)$ as functions of $\bar x$ could easily be constructed.

Table 2 summarizes the results of applying these estimation procedures to the four distributions in Table 1, and also shows χ^2 values based upon comparison of the observed frequencies with those calculated from insertion of the estimated parameter in equation (2). The χ^2 values for the eggs give no sign of appreciable deviation from the hypothesis of Poisson distributions. The fact that the gall-cell distributions also yield low values of χ^2 shows the hypothesis that egg mortality is independent of x to be not contradicted by the data to any appreciable extent. Both $\hat\lambda$ and $\hat\mu$ agree remarkably closely in the two years.

Summary of estimates of Poisson parameters, variances, and homogeneity tests

	1935	1936
$\bar{x}$	3.020	3.034
$\hat{\lambda}$	2.845	2.860
$V(\hat{\lambda})$	0.0220	0.0371
χ^2	4.17 (4 d.f.)	6.83 (4 d.f.)
$\bar{y}$	2.283	2.336
$\hat{\mu}$	1.962	2.028
$V(\hat{\mu})$	0.0028	0.0088
χ^2	6.89 (4 d.f.)	4.87 (4 d.f.)

In the calculation of χ^2, frequencies of flower-heads with 6 or more eggs or gall-cells were combined.

From equation (1), the maximum likelihood estimate of the mortality rate is

$$\hat{\theta} = 1 - \frac{\hat{\mu}}{\hat{\lambda}}. \tag{5}$$

Here, as in most practical situations, $\hat{\lambda}$ is much larger than its standard error, and the asymptotic variance of $\hat{\theta}$ can be safely used:

$$V(\hat{\theta}) = [V(\hat{\mu}) + (1 - \hat{\theta})^2 V(\hat{\lambda})] \div \hat{\lambda}^2. \tag{6}$$

Hence, in 1935

$$\hat{\theta} = 0.310 \pm 0.040$$

and in 1936

$$\hat{\theta} = 0.291 \pm 0.058.$$

The estimates are very close to Varley's, but the standard errors are larger because of the allowance that has now been made for sampling errors in the distribution of numbers of eggs.

4. *Plackett's Method*

An elegant alternative to maximum likelihood estimation for the parameter of a truncated Poisson distribution is due to Plackett (1953). He showed that

$$\lambda^* = \sum_{x=2}^{\infty} x n_x / N, \tag{7}$$

n_x being the number of observations for the value x, is an unbiased estimator of λ whose efficiency never falls below 95%. The estimator

may alternatively be written

$$\lambda^* = \bar{x} - \frac{n_1}{N}.$$ (7.1)

He further showed that

$$V(\lambda^*) = (N\lambda^* + 2n_2)/N^2$$ (8)

is an unbiased estimator of the variance of λ^*. The numerical estimates of λ and μ obtained in this way are almost the same as those in Table 3, and the mortality rates are estimated as

$$\theta^* = 0.306 \pm 0.042$$

in 1935 and

$$\theta^* = 0.273 \pm 0.061$$

in 1936. Thus estimates with only slightly larger standard errors, and standard errors that are now known to be valid even in small samples, are obtained with much less labour.

5. *Discussion*

Varley's observations on the numbers of eggs per flower-head are in close agreement with what would be expected if the number of eggs per flower-head followed a Poisson distribution and the proportion of eggs failing to produce gall-cells were independent of the number of eggs per flower-head. This is no demonstration that competition within the flower-head is completely absent: doubtless competition and an increased death rate of eggs must occur if the number of eggs per flower-head is substantially increased, since a flower-head cannot hold a gall with more than about 15 gall-cells, but in samples of the size taken in 1935 and 1936 the effect lay within the limits of error. This re-analysis modifies Varley's previous conclusions (*loc. cit.* p. 175), for the available information is in fact insufficient to demonstrate an increase of larval mortality with increasing size of egg batch. It would indeed be interesting to know more about the sensitivity of the tests used here, and in particular to know how large an effect of competition could escape detection in this number of observations, but that appears to involve a much more difficult analysis. One flaw in both the present analysis and that used previously by Varley is that the χ^2 tests relate to any form of heterogeneity in the mortality rate for different numbers of eggs per flower-head. The real interest lies, however, in the possibility that the mortality rate shows a regular trend as x increases; one would therefore like to be able to isolate a single

degree of freedom in the appropriate χ^2 that would represent a regression of mortality rate on x, but unfortunately there appear to be substantial difficulties in thus complicating the analysis. The values of χ^2 in Table 2 are in fact so small as to leave little hope that such a component would be statistically significant, and inspection of the observed and expected frequencies supports this opinion. Nothing can be said about egg mortality at higher densities, but within the range of the numbers per flower-head observed an average of about 30% or slightly less seems appropriate for both years.

It may be objected that the validity of the estimate rests upon the assumption of a Poisson distribution for the egg frequencies, an assumption for which there is no theoretical basis. Analogous assumptions are of course implicit in many techniques for estimating a population characteristic from a sample, and the justification lies in the choice of a technique that makes the estimate relatively insensitive to the exact terms of the assumption. That no significant deviation from the Poisson is found does not prove that the distribution was Poisson, any more than the comparison of the egg and gall-cell frequencies has proved that the mortality is constant. The Poisson distribution is introduced as a convenient simple model that is not contradicted by the data; since the purpose of the analysis is only to examine the ratio of comparable means for the two frequency distributions, the precise algebraic formulation used is not very important. Any attempt to examine the mortality between oviposition and gall-cell formation without specification of a model for the egg distribution is doomed to failure because each egg frequency must be represented by a separate parameter. Varley's analysis in fact assumed that the true egg distribution was exactly proportional to that observed, though one would scarcely seriously maintain that in 1935 nearly 0.7% of flower-heads had exactly 12 eggs while none had 10, 11, or 13!. All that is required for the estimation of the mortality rate is an estimate of the number of eggs per flower-head in a particular group of flowers and an estimate of the number of gall-cells per flower-head in the same or an exactly corresponding group of flowers: the whole difficulty lies with the empty flower-heads, since some that contain eggs will be empty in respect of gall-cells. The Poisson distribution involves only one new parameter, and is about the simplest assumption possible on which eggs and gall-cells can be averaged over comparable groups of flowers; it has been shown to fit the data excellently in both years. Since the hypothesis of a Poisson distribution and a constant mortality is not contradicted, that of a constant mortality without specification of distribution is tenable under the conditions studied. An alternative might give very

different estimates of the mean numbers of eggs and gall-cells, because of a different relationship between the (unobservable) zero class and the others, but, if it were to fit the non-zero observations equally well, it would almost certainly give much the same ratio of the means. If an alternative such as the negative binomial (Sampford, 1954) were employed, however, the information provided by the sample would have to be used in estimating a larger number of parameters,* so that larger standard errors and a less sensitive test of the hypothesis of random mortality would probably be obtained.

6. *Summary*

Records of the frequency distribution of eggs and gall-cells of the knapweed gall-fly in flower-heads of black knapweed have been used in illustration of the use of the truncated Poisson distribution for representing biological observations. If mortality between oviposition and gall-cell formation were entirely random and did not depend upon the density of eggs in a flower-head, a truncated Poisson distribution of eggs would lead to a similar distribution of cells. The analysis shows that, in both years of recording, the observations on eggs and cells were excellently described by distributions of this type, so that no evidence against random mortality appears. It is unlikely that any different conclusion would be reached by using any alternative formulation of the egg distribution. However, nothing is known of the power of the test for detecting deviations from random mortality.

For both the 1935 and the 1936 data, the estimated mortality rate is about 30%, but the samples are not large enough to determine this very exactly: even if the information from the two years is pooled, a standard error of about $3\frac{1}{2}\%$ must be attached to this estimate.

---

*In addition, new technical difficulties might be introduced by the cell distribution assuming a much more complicated form than the egg distribution (though this does not occur for the negative binomial).

REFERENCES

Cohen, A. C. (1954). Estimation of the Poisson parameter from truncated samples and from censored samples. *Journal of the American Statistical Association, 49,* 158–68.

David, F. N., and Johnson, N. L. (1952). The truncated Poisson. *Biometrics, 8,* 275–85.

Plackett, R. L. (1953). The truncated Poisson distribution. *Biometrics, 9,* 485–8.

Rider, P. R. (1953). Truncated Poisson distributions. *Journal of the American Statistical Association, 48,* 826–30.

Sampford, M. R. (1954). The truncated negative binomial distribution. *Biometrika, 42*: 58–69.

Varley, G. C. (1947). The natural control of population balance in the knapweed gall-fly (*Urophora jaceana*). *Journal of Animal Ecology, 16,* 139–87.

QUERY: I have an experiment with several lots with different
117 numbers of animals. There is strong evidence (including Bart-
lett's test) that the variances in the lots are different. The experi-
ment is similar to that in Snedecor's text (4th edition) section 10.8. He
does not give a test for the means if the variances differ. I should think
some kind of weighted analysis of variance could be used. Is there any
method available for making the analysis?

ANSWER: As there is apparently strong evidence that heterogeneity
of variance is present among the lots it seems advisable, as
the inquiry suggests, to use a test which would allow for
such heterogeneity rather than the ordinary analysis of variance test
which is based on the assumption of a common variance for all the lots.

It is well-known that the usual F-statistic for testing the null hy-
pothesis of equality of lot means is

$$V_1 = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{\sum_{i=1}^{k}\sum_{j=1}^{k}(x_{ij} - \bar{x}_i)^2} \frac{n - k}{k - 1}$$

which has, under the assumption of common variance and the null
hypothesis, an F distribution with degrees of freedom $k - 1$, $n - k$. In
the experiment referred to in Snedecor's book (see Table 10.12), $k = 8$,
$n = 56$; the lot sizes n_i and the lot means $\bar{x}_i$ are also given there. It
turns out that $V_1 = 2.97$ which is just short of the 1% point 3.04 but
well above the 5% point 2.21.

In a recent paper Box [1] has investigated the distortion in the dis-
tribution of V_1 when the assumption of equality of lot variances is
violated. In Table 4 (loc. cit), for instance, he notes that for certain
specified lot sizes and lot variances the true level of a 5% F-test which
employs V_1 may be much less or much greater than 5%, depending on
the particular lot sizes and variances. He also suggests as a working
approximation to the distribution of V_1 under the null hypothesis of
equality of means to regard V_1/b as having an F-distribution with de-
grees of freedom h', h where

$$b = \frac{n - k}{n(k - 1)} \cdot \frac{\sum_{i=1}^{k} (n - n_i)\sigma_i^2}{\sum_{i=1}^{k} (n_i - 1)\sigma_i^2}$$

$$h' = \frac{\left\{\sum_{i=1}^{k} (n - n_i)\sigma_i^2\right\}^2}{\left\{\sum_{i=1}^{k} n_i\sigma_i^2\right\}^2 + n \sum_{i=1}^{k} (n - 2n_i)\sigma_i^4}$$

$$h = \frac{\left\{\sum_{i=1}^{k} (n_i - 1)\sigma_i^2\right\}^2}{\sum_{i=1}^{k} (n_i - 1)\sigma_i^4}$$

Here σ_i^2 is the true variance of the normal distribution corresponding to the ith lot. It is, of course, presumptuous to apply the above approximation with the sample variance s_i^2 replacing σ_i^2. If it were applied to the data of Table 10.8 it would be impaired even further by the fact that the lot sizes are small (4 to 10). The technique is stated here, however, in the belief an experimenter may find it useful if he has a sample which involves large lot sizes.

From the data it is found that

$$b = 0.8952, \qquad h' = 4.5, \qquad h = 27.7, \qquad \frac{V_1}{b} = 3.32.$$

By interpolation in the F-tables it can be seen that the probability of exceeding the value 3.32 is approximately 3%. Application of the ordinary F test (ignoring inequality of variances) yielded a value just short of the 1% point; however the approximation is too rough here to certify a contradiction.

An apparently more promising method of coping with inequality of lot variances is to use the statistic

$$V_2 = \sum_{i=1}^{k} w_i(\bar{x}_i - \hat{\mu})^2$$

where

$$\hat{\mu} = \frac{\sum_{i=1}^{k} w_i\bar{x}_i}{\sum_{i=1}^{k} w_i}$$

and

$$w_i = \frac{n_i}{s_i^2}$$

Under the null hypothesis of equality of means, V_2 is distributed approximately as χ^2 with $k - 1$ degrees of freedom for large values of n_i. A rationale for using V_2 is that when σ_i^2 replaces s_i^2, the test is equivalent to the likelihood ratio test for equality of means when the variances are known (and need not be equal).

Welch [2] has improved the χ^2 approximation to V_2 by dividing it by a correction factor. Specifically, he recommends using the statistic

$$\frac{V_2}{1 + \dfrac{2(k-2)}{k^2 - 1} \sum\limits_{i=1}^{k} \dfrac{1}{n_i - 1}\left(1 - \dfrac{w_i}{\sum\limits_{i=1}^{k} w_i}\right)^2} = V_3, \qquad \text{say}$$

as if it were distributed as F with degrees of freedom $\hat{f}_1, \hat{f}_2$, where $\hat{f}_1 = k - 1$

$$\hat{f}_2 = \left[\frac{3}{k^2 - 1} \sum_{i=1}^{k} \frac{1}{n_i - 1}\left(1 - \frac{w_i}{\sum\limits_{i=1}^{k} w_i}\right)^2\right]^{-1}$$

For the data of table 10.8 it turns out that

$$\hat{\mu} = 2.89, \qquad \hat{f}_1 = 7, \qquad \hat{f}_2 = 16.6, \qquad V_3 = 4.53$$

By linear interpolation in the F-tables it is seen that the 1% and $\frac{1}{2}$% points are approximately 4.00 and 4.66 respectively. It will be noted that the value of V_3 falls short of the $\frac{1}{2}$% point. It should also be pointed out that due to the method by which Welch developed this approximate distribution he can state only that it holds to order $1/(n_i - 1)$. Since the lot sizes in the data range from 4 to 10 it is apparent that the approximation is too rough to be trustworthy for the example considered.

James [3] has also developed an approximation to the distribution of V_2 which holds to order $1/(n_i - 1)$. Specifically, for a given level of significance P, say, he finds a function $h(w_1, w_2, \cdots, w_k)$ such that

$$P\{V_2 > h(w_1, w_2, \cdots, w_k)\} = P$$

For an approximation to the order $1/(n_i - 1)$ he recommends setting

$$h(w_1, w_2, \cdots, w_k)$$
$$= \chi_P^2\left[1 + \frac{3\chi_P^2 + k + 1}{2(k^2 - 1)} \sum_{i=1}^{k} \frac{1}{n_i - 1}\left(1 - \frac{w_i}{\sum\limits_{i=1}^{k} w_i}\right)^2\right]$$

where χ_P^2 is the value of χ^2 with $k - 1$ degrees of freedom which is exceeded with probability P. As pointed out above, the small lot sizes of the example cause such an approximation to be rough. The value of h

computed for the data turns out to be 34.49 when P is .005. Since the computed value of V_2 is 39.36, a test based on this approximate distribution would recommend rejection of the null hypothesis at the $\frac{1}{2}\%$ level. As a consequence of the small lot sizes it is not at all surprising that this contradicts the decision not to reject at the $\frac{1}{2}\%$ level based on Welch's approximation.

It may also be remarked that James (loc cit.) has developed a more refined approximation to the distribution of V_2 which is of order $(1/(n_i - 1))^2$, but the computations involved are so exceedingly tedious as to discourage its use.

In conclusion, this author's answer to the query as to whether there is available some kind of weighting technique for analyzing an experiment such as that of Table 10.12 in which heterogeneity of variance is present, would be a qualified yes. Yes if the lot sizes are not small, in which case one could use the results of Welch or James. What is meant here by "not small" cannot, at this stage of available results, be categorically defined since it is merely the order of the approximation which is $1/(n_i - 1)$. Thus, for instance, if all the lots sizes exceed 10 the error which results in using the given approximation would be of an order not exceeding .01.

A final word of warning might be added here. Since distortions due to inequality of variances alter the effective level of significance of the ordinary analysis of variance test, a policy of using the F-test and ignoring variance heterogeneity may fortuitously lead to a correct decision in some circumstances. For instance, if the lot sizes and variances are such that the effective level of 5% F-test is increased to 10 or 15%, there will tend to be an increase of power; consequently, an ordinary F-test, as a result of such distortion would tend to detect the falsehood of the null hypothesis more frequently than it would without distortion. However, since the effective level might also decrease as a result of distortion, a policy of ignoring variance heterogeneity is self-contradictory.

REFERENCES

[1] G. E. P. Box, "Some theorems in quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification," *Annals of Mathematical Statistics*, *25*, (1954), pp. 290–316.
[2] B. L. Welch, "On the comparison of several mean values: An alternative approach," *Biometrika*, *38*, (1951), pp. 330–336.
[3] G. S. James, "The comparison of several groups of observations when the ratios of the population variances are unknown," *Biometrika*, *38*, (1951), pp. 324–329.

JOHN GURLAND

The Biometric Society

BIOMETRICS

TABLE OF CONTENTS

PARTIALLY REPLICATED LATIN SQUARES

W. J. Youden* and J. S. Hunter**

Institute of Statistics, N. C. State College
Raleigh, N. C.

Latin squares have been used for over 30 years in agricultural field trials (1). Often the double restriction imposed by the row and column arrangement brought a welcome reduction in the mean square for error. The row and column mean squares were of no interest. Eventually the Latin square arrangement found application where the rows and columns corresponded to clearly defined physical entities. Youden (3) in studies on tobacco mosaic virus observed that the number of lesions produced on leaves depended upon the position of the leaf on the plant and that plants also differed markedly in susceptibility to lesions. Furthermore, for a given lot of plants, the effect of leaf position was closely the same from one plant to another. Very large reductions in the error mean square resulted from the use of a Latin square. The plants were columns and leaf positions were rows. Here there was some interest in the mean squares for rows and columns though the chief concern was with the treatments applied to the leaves. Later Yates (2) introduced confounding into Latin square designs.

Probably it was inevitable that the Latin square arrangement would be tried when the rows and columns were used for factors that not only were likely to interact with the treatments (letters) but also with each other. The form or appearance of a Latin square remains but the substance is lost. The fractional replication of a factorial experiment that results from this practice is a particularly unfortunate one as all

main effects are directly confounded with two factor interactions and the residual (in 3×3 squares or larger) becomes a snare for the unwary. The plausible appearance of getting something for nothing has trapped and will continue to trap novices in the use of experimental design. For this reason there may be some advantage in making available a slight extension of the Latin square that will give an indication as to whether or not the usual requirement of additive effects for rows, columns and treatments has been met.

The proposed extension consists in performing for a $k \times k$ Latin square k additional experiments so chosen that each row, column and letter enters into $k + 1$ measurements. For convenience the duplicated cells may be shown lying along a principal diagonal. Randomizing the rows and columns will change the pattern of duplicated cells.

	c_1	c_2	c_3
r_1	A	B	CC'
r_2	C	AA'	B
r_3	BB'	C	A

	c_1	c_2	c_3	c_4	c_5	c_6
r_1	AA'	C	F	E	D	B
r_2	D	BB'	E	F	C	A
r_3	B	D	CC'	A	F	E
r_4	E	F	A	DD'	B	C
r_5	F	A	B	C	EE'	D
r_6	C	E	D	B	A	FF'

	c_1	c_2	c_3	c_4
r_1	AA'	C	B	D
r_2	B	D	A	CC'
r_3	D	BB'	C	A
r_4	C	A	DD'	B

It is not difficult to construct various arrangements for each size Latin square. Enough symmetry is retained to lead to reasonably convenient sets of estimates for row, column and treatment effects. The analysis of variance for a $k \times k$ square follows:

TABLE 1

Analysis of variance for partially replicated Latin squares.

Item	d.f.	Sum of squares
Rows	$(k-1)$	$\sum_{i=1}^{k} \dfrac{R_i^2}{k+1} - \dfrac{G^2}{k(k+1)}$
Cols. adjusted for rows	$(k-1)$	$\dfrac{k+1}{k(k+2)} \sum_{i=1}^{k} (C_i')^2$
Treats. adj. for rows, cols.	$(k-1)$	$\dfrac{k+2}{k(k+3)} \sum_{i=1}^{k} (T_i'')^2$
"Interaction"	$(k-1)(k-2)$	By difference
Error (duplicates)	k	$\dfrac{1}{2} \sum \left(\begin{array}{c}\text{differences}\\ \text{of duplicates}\end{array}\right)^2$
Total sum of squares	$k(k+1) - 1$	$\sum_{i,j=1}^{k} y_{ij}^2 - \dfrac{G^2}{k(k+1)}$

y_{ij} = an observation in the ith row, jth column

G = Grand Total $\qquad$ R_i = Total of ith row

C_i = Total of ith column $\qquad$ T_i = Total of ith treatment

$$C_i' = C_i - \frac{R+G}{k+1}$$

where R is that row total associated with column i in which the duplicate occurs.

$$\sum_{i=1}^{k} C_i' = 0$$

$$T_i'' = T_i - \frac{R+C+G}{k+2}$$

where R and C are the row and column totals for row and column in which treatment i is duplicated.

$$\sum_{i=1}^{k} T_i'' = 0$$

Adjusted treatment = mean =

$$t_i = \frac{(k+2)}{k(k+3)} T_i'' + \frac{6}{k(k+1)}$$

Variance difference between treatment means

$$= V(t_i - t_j) = 2\,\frac{(k+2)}{k(k+1)}\,s^2$$

As an example consider some data obtained on the density of small bricks. Three different sizes of powder particles were used. Each size was compressed under three pressures and fired at three temperatures. Finally the 27 combinations of this $3 \times 3 \times 3$ factorial were run in duplicate. These data are given in full and then portions of them used to show what happens when a subset of 9 in the form of a Latin square are chosen on the presumption that useful information will be obtained. The same 9 values will then be supplemented by the three duplicates required for the partially replicated Latin square.

The complete set of 54 measurements is given in Table 2. The values in italics (using only the first when both are in italics) are those used in a Latin square selection. The analysis of variance for these 9 results is listed in Table 3 alongside the mean squares for the complete set of data. It is abundantly clear that no useful interpretation is possible using the 9 results nor is there any way to ascertain that the error variance is, in fact, a good deal smaller than the residual mean square of 2258.

<div align="center">TABLE 2</div>

Densities of Briquettes formed from three sizes of particles, compressed at three pressures and fired at three temperatures. The decimal points are omitted. Duplicate values are separated by commas.

Size	Pressure	Temperature, degrees Fahrenheit		
		1900	2000	2300
5–10	5.0	945, *961	*933*, 968	962, *950
	12.5	969, 960	944, 882	*942, 958*
	20.0	*964*, *964	949, 964	965, *974
10–15	5.0	905, 897	969, 927	*908*, 892
	12.5	*936*, 946	925, 985	892, 904
	20.0	940, 924	*905*, *943*	950, 917
15–20	5.0	*842*, *845	848, 872	851, *881
	12.5	868, 790	*981*, 989	872, 879
	20.0	845, *880	993, 1020	*890*, *902

TABLE 3
Analysis of Variance for Density Data.

Item	All 54 results		Latin square 9 results	
	d.f.	M.S.	d.f.	M.S.
Temp.	2	6011	2	676
Size	2	17065	2	1404
Pres.	2	3946	2	2598
$T \times S$	4	6834		
$T \times P$	4	466	2	2258
$S \times P$	4	1798		
$T \times S \times P$	8	2170		
Dupli.	27	456		

TABLE 4
Analysis of variance for partially replicated Latin square as explained in Table 1. Data used are the italicized values in Table 2.

Item	d.f	S.S.	M.S.
Size	2	7175.166	
Temp. corr. for size	2	2737.233	
Pres. corr. for size and temp.	2	6662.600	3331
"Interaction"	2	4083.416	2042
Error	3	854.500	285
Total	11	21512.916	

If the experimenter wants to take a chance that the interactions are small the experiment should furnish the means of demonstrating that the gamble has been won. To this end the other three values underlined in Table 2 are combined with the 9 values and the 12 results examined by the formulas in Table 1. Table 4 shows the analysis of variance, the pressures taking the role of treatments. The three additional measurements give a valid estimate of the experimental error provided the order was suitably randomized. The small value for the error variance relative to the interaction mean square gives warning that the interactions are not negligible and that the main effects may not be what they seem to be. It is of course recognized that not all possible interaction effects are measured by this technique, but that these interactions are in fact sampled by the degrees of freedom available.

The sum of squares for temperature when corrected for size and pressure is 2273.4. The sum of squares for size, when corrected for temperature and pressure, is 3554.6. These will be needed provided the duplicates give assurance of little or no interaction. The computation of these corrected sums of squares may be obtained by permuting the roles of R_i, C_i and T_i in the formulas shown in Table 1.

There is another alternative to the one third replicate of the $3 \times 3 \times 3$ factorial if there exists any strong misgivings about the absence of interactions. In Table 2 eight of the entries are marked with an asterisk. These entries constitute a $2 \times 2 \times 2$ experiment using the lowest and highest temperatures, smallest and largest particles, and the lowest and highest pressures. The analysis of variance for these eight results is shown in Table 5.

TABLE 5
Analysis of variance for $2 \times 2 \times 2$ factorial.

Item	d.f.	M.S.	
Size	1	20808	
Temperature	1	648	
Pressure	1	512	
$T \times S$	1	162	
$T \times P$	1	50	
$S \times P$	1	50	
$T \times S \times P$	1	338	

With all the limitations of single degrees of freedom the experimenter can feel fairly confident that particle size is important. The experiment fails to reveal interactions that appear to arise with the intermediate levels of the factors that were omitted. Altogether the example shows that skimping of the measurements cuts down on the information. Experimenters know that small or even moderate effects are likely to be missed when only a few measurements are made. Furthermore the use of a Latin square arrangement gave results that are likely to be dangerously misleading to workers with little experience in the design and analysis of experiments. For that matter, even the analysis for the complete experiment may mislead the beginner. The small mean square for $T \times P$ may lead to the hasty conclusion that these factors do not interact. But if the three temperature-pressure tables are examined separately for each size of particle, significant interactions will be found. In the analysis of the whole set of data these interactions are compen-

sating. The large mean square for the three factor interaction gives warning of this possibility.

In summary, the experimenter is not always aware that additivity of rows, columns and treatments is a basic assumption for the Latin square. The experimenter sees only that, by identifying rows, columns, and letters with experimental factors, a small subset of treatments is specified. Ultimately the experimenter may learn that there is no unambiguous interpretation of these so called Latin squares unless he has information about the experimental error. The slightly replicated Latin square directs attention to the need for this estimate of error. The degrees of freedom for error are few. On the other hand the duplicates have been chosen to facilitate the examination of the data.

REFERENCES

1. Fisher, R. A., *Statistical Methods for Research Workers*. 1st ed. 1925 Oliver and Boyd. Edinburgh, England.
2. Yates, F., The Design and Analysis of Factorial Experiments. *Techn. Comm. No. 35*, 1935. *Imper. Bur. of Soil Science*, Harpenden, England.
3. Youden, W. J. and H. P. Beale, A statistical study of the local lesion method for estimating tobacco mosaic virus. *Contributions from Boyce Thompson Institute 6*, 437. 1934.

THE RELATIVE SIZE OF THE INTER- AND INTRA-BLOCK ERROR IN AN INCOMPLETE BLOCK DESIGN*

W. A. THOMPSON, JR.

Virginia Polytechnic Institute

1. INTRODUCTION

Results of scientific experiments are frequently classified according to the factors contributing to the observed data. Thus, several technicians may test a property of a certain type of shoe on several different walking courses, and y_{ij} is the score given to the shoe by the ith technician on the jth course. The shoe to shoe factor for a given type shoe is assumed negligible. The two factors considered in this example are then (i) different technicians and (ii) different courses.

An individual technician's gradings will not be completely reproducible, i.e., if he repeatedly grades the same type shoe on the same course he will not always give it the same score. He will tend to give scores which fall in a regular manner about the actual value for the particular shoe and course. This variation is then the error contribution to the score.

If, as is usually the case, we wish to draw inferences about locations other than those where we perform our experiments, it may be more reasonable to assume that the course effect is composed of a large number of effects which have a tendency to counterbalance each other. This suggests to the author that we should also assume the course effects to be random ones, and we will then have two sources of variability with which to deal: (i) the variability in performance due to the different courses and (ii) the variability due to all other factors which we have called the error variability.

Hence, if the ith technician works on the jth course we will assume that

$$y_{ij} = \alpha + \tau_i + \beta_j + \epsilon_{ij}$$

α being an average effect over all technicians and courses, τ_i the effect of the ith technician and β_j the effect of the jth course; ϵ_{ij} is the error contribution. The β_j's are assumed to be independent observations from a normal distribution with mean zero and variance σ_1^2 ; the ϵ_{ij}'s are assumed to be independent and identically distributed with mean zero and variance σ^2.

The reader may recognize that we are dealing with a "mixed" model since the β's are assumed to be random and the τ's fixed.

We now wish to investigate the size of σ_1^2 relative to σ^2. We do this by testing hypotheses concerning σ_1^2/σ^2 ($= \mu$) or finding a confidence interval for this ratio. It is well known that for a randomized block design this can be done by using the F table. The purpose of this paper is to indicate the further refinements necessary for handling an incomplete block design. The method used here was proposed by Wald (5) and developed by the author (4).

In Section 2, in order to present the reader with a concrete illustration of this method, we shall enlarge upon the shoe example. In Section 3, we shall present, for the use of the experimenter, a general set of operating rules. In Section 4 we provide an illustration of the use of the rules given in Section 3 and discuss the possibility of designing experiments with respect to blocks as well as with respect to treatments.

2. EXAMPLE

The illustration which we shall carry over from 1 is furnished by a military experiment in which we test a type of shoe in order to determine its desirability under various walking conditions.

2.1 *Background and Data for the Illustration.*

In this example a number of technicians are asked to rate a particular kind of shoe according to the following scale:

1. Extremely unsatisfactory
2. Very unsatisfactory
3. Moderately unsatisfactory
4. Slightly unsatisfactory
5. Not good, not bad
6. Slightly satisfactory

7. Moderately satisfactory
8. Very satisfactory
9. Extremely satisfactory

Thus, the technicians will give the higher grades to the shoes which they feel are more desirable.

Let us suppose that in this experiment there are six different courses on which the technicians will judge the shoes; these courses are assumed to be randomly chosen from the locations at which it is likely that the shoes will be used.

If we use twenty technicians and six courses, we could let each technician go over each course. We would then have 120 different ratings or y_{ij}'s. However, it may be too expensive or inconvenient to get as many as 120 different ratings, and 60 may be the largest number of ratings that is feasible. Of course, the more ratings that are made, the more accurate will be the knowledge gained from the experiment. On the other hand, there are other ways besides enlarging the number of ratings which will increase the information obtained; particularly, the design of the experiment will insure that resources are used to the best advantage.

In the case at hand a very worthwhile design requiring 60 different ratings would be the following one

course \ technicians	1 5 9 13 17	2 6 10 14 18	3 7 11 15 19	4 8 12 16 20
1	X	X		
2	X		X	
3	X			X
4			X	X
5		X		X
6		X	X	

That is, technicians 1, 5, 9, 13 and 17 go over courses 1, 2, 3; technicians 2, 6, 10, 14 and 18 go over courses 1, 5 and 6, etc. [This is design SR6 Bose, Clatworthy, and Shrikhande (2).]

Using the method of scoring and the design indicated above, the following table contains the raw scores of the actual experiment.

technician \ course	1	2	3	4	5	6	Total	Mean
1	3	8	9				20	6.67
2	1				1	5	7	2.33
3		9		4		4	17	5.67
4			9	5	3		17	5.67
5	3	6	6				15	5
6	4				7	1	12	4
7		2		3		9	14	4.67
8			5	6	9		20	6.67
9	6	2	9				17	5.67
10	4				5	9	18	6
11		6		8		1	15	5
12			9	5	8		22	7.33
13	6	5	1				12	4
14	2				6	6	14	4.67
15		4		8		7	19	6.33
16			9	5	2		16	5.33
17	5	9	7				21	7
18	5				1	3	9	3
19		9		6		7	22	7.33
20			6	7	9		22	7.33
Total	39	60	70	57	51	52	329	

Thus, technician 1 rated the shoe to be moderately unsatisfactory on course 1, very satisfactory on course 2 and extremely satisfactory on course 3.

A property of this design which is worth noting is that a number of the technicians do exactly the same things. This makes it possible to instruct and to transport the men in groups and so carry out the experiment with a minimum of confusion. The advantage of transporting the technicians in groups becomes apparent when one alters the circumstances of the present problem and considers the possibility of having 4 groups of 5 observers each watch six different maneuvers.

We have indicated that we are interested in determining the size of

$$\mu = \sigma_1^2/\sigma^2 = \frac{\text{variation due to courses}}{\text{variation due to error}}.$$

In order to show why we might be interested in this information let us outline some of the possible actions which we might take as a result of knowing σ_1^2/σ^2 exactly.

If the course variability is relatively small, then some of the possible

actions which might result as a consequence of having run the experiment are:

1) In analysing the results of the experiment we will "pool" the data over all courses to determine how satisfactory the shoe being tested is. That is, we treat the data as though it was taken all on the same course instead of on six different ones.

2) If the shoe being tested has an overall grade of satisfactory, then we may recommend that the army use this shoe in all locations represented in the test. Hence we "standardize" on this particular shoe.

If, however, we find σ_1^2/σ^2 large, then we will be unable to standardize on a particular shoe since our experiment will have indicated that the desirability of the shoe is not constant over the locations of expected use. A re-examination of the testing procedure and investigation of the properties of the shoe may be indicated.

2.2 *Some Reasons.*

The reader may have a better insight for what is going on if we sketch very briefly some of the reasons for analysing the experiment as we shall in the next section.

In order to remove the effect of the different technicians we average them out by taking the adjusted course totals.

2.2.1 p_i = adjusted course total for course i
            = ith course total—sum of means of technicians who are judging on the ith course.

In some designs and for fixed β's the p's would be estimates of the course effects, but in the present one for given β's

$$E(p_1) = \frac{20}{3}\,\beta_1 - \frac{5}{3}\,(\beta_2 + \beta_3 + \beta_5 + \beta_6)$$

$$E(p_2) = \frac{20}{3}\,\beta_2 - \frac{5}{3}\,(\beta_1 + \beta_3 + \beta_4 + \beta_6)$$

$$E(p_3) = \frac{20}{3}\,\beta_3 - \frac{5}{3}\,(\beta_1 + \beta_4 + \beta_2 + \beta_5)$$

2.2.2

$$E(p_4) = \frac{20}{3}\,\beta_4 - \frac{5}{3}\,(\beta_2 + \beta_5 + \beta_3 + \beta_6)$$

$$E(p_5) = \frac{20}{3}\,\beta_5 - \frac{5}{3}\,(\beta_1 + \beta_4 + \beta_3 + \beta_6)$$

$$E(p_6) = \frac{20}{3}\,\beta_6 - \frac{5}{3}\,(\beta_1 + \beta_2 + \beta_4 + \beta_5)$$

or $E(p) = D\beta$ where β, is the ith course effect and D is the matrix of coefficients of the β's.

The unconditional variance-covariance matrix of p_1, p_2, $\cdots$, p_6 is $D\sigma^2 + D^2\sigma_1^2$.

At this point we choose a set of orthogonal linear functions of p_1, $\cdots$, p_6, say z_1, z_2, $\cdots$, z_6. z_1, z_2, $\cdots$, z_6 may be chosen so that z_1, z_2, $\cdots$, z_5 have the distribution:

$$\text{constant exp}\left(-\frac{1}{2}\sum_{i=1}^{5}\frac{z_i^2}{e_i\sigma^2 + e_i^2\sigma_1^2}\right) \qquad dz_1 \cdots dz_5$$

while z_6 is zero with probability one. For the mathematical reader e_1, $\cdots$, e_5 are the non-zero characteristic roots of D.

We now consider the ratio

$$\left(\frac{1}{5}\sum_{i=1}^{5}\frac{z_i^2}{e_i\sigma^2 + e_i^2\sigma_1^2}\right)\Big/\frac{1}{\sigma^2} \qquad \text{(M.S.E.)}$$

where, of course, M.S.E. stands for mean square due to error.

We may prove that the numerator and denominator of this ratio are independently distributed according to the Chi-Square distribution, and hence the ratio has the F distribution. We denote this ratio by $F(\mu)$ where $\mu = \sigma_1^2/\sigma^2$.

2.2.3
$$F(\mu) = \frac{1}{5(\text{M.S.E.})}\sum_{i=1}^{5}\frac{z_i^2}{e_i + e_i^2\mu}.$$

Wald (5) points out that $F(\mu)$ is a decreasing function of μ, and hence, if

$$\Pr[F(\mu) < F_{1-\alpha}] = 1 - \alpha,$$

then

2.2.4
$$\Pr(\mu > \mu_{1-\alpha}) = 1 - \alpha,$$

where $\mu_{1-\alpha}$ is defined by $F(\mu_{1-\alpha}) = F_{1-\alpha}$ and where $F_{1-\alpha}$ is the ordinate of the cumulative distribution function of an F statistic with the appropriate number of degrees of freedom. We will then have placed an upper confidence bound on μ with a probability of $1 - \alpha$. We may also find a lower confidence bound for μ and hence a confidence interval for μ (though not a unique one).

Thus, in order to make confidence statements about $\mu = \sigma_1^2/\sigma^2$, we must solve equations of the type

$$\sum_{i=1}^{5}\frac{z_i^2}{e_i + e_i^2\mu} = a,$$

where $a = F_{1-\alpha}$ (M.S.E.)5. For the present incomplete block design it is true that there are only two different e's (i.e., the matrix D has only two different characteristic roots). We thus have a much easier equation to solve:

2.2.5
$$\frac{\sum_1}{e_1 + e_1^2\mu} + \frac{\sum_2}{e_2 + e_2^2\mu} = a,$$

where

$$\sum_1 = \sum_{\substack{i \\ e_i = e_1}} z_i^2 = \frac{e_1}{e_1 - e_2}\left(\sum_{i=1}^{6} p_i^2 - e_2 \sum_{i=1}^{6} m_i p_i\right)$$

and

$$\sum_2 = \sum_{\substack{i \\ e_i = e_2}} z_i^2 = \frac{e_2}{e_2 - e_1}\left(\sum_{i=1}^{6} p_i^2 - e_1 \sum_{i=1}^{6} m_i p_i\right).$$

Here m_1, m_2, $\cdots$, m_6 are found by solving equations 2.2.2 for the β's after having substituted p_i for $E(p_i)$. The evaluation of $\sum_1$ and $\sum_2$ in terms of $p_1 \cdots p_6$ and $m_1 \cdots m_6$ is not immediately evident, but will not be discussed here.

We may now solve equation 2.2.5 by clearing of fractions, evaluating $\sum_1$ and $\sum_2$ in terms of $p_1, \cdots, p_6$ and considering the solutions of the resulting quadratic equations. The non-extraneous solution to the quadratic is then also a solution to our original equation; it turns out to be

2.2.6
$$\mu_0 = \frac{-d + \sqrt{d^2 - 4bc}}{2b}$$

where

2.2.7
$$\Delta^* = e_1 e_2$$
$$H^* = (e_1 + e_2)$$
$$a = F_{1-\alpha}(\text{M.S.E.})5$$
$$b = a e_1 e_2 = a\Delta^*$$
$$c = a - \sum_{i=1}^{6} m_i p_i$$
$$d = c(e_1 + e_2) + \sum_{i=1}^{6} p_i^2$$
$$= cH^* + \sum p_i^2$$

2.3 *The Analysis.*

We will use the shoe data that we have been discussing to illustrate two methods of studying the ratio

$$\mu = \frac{\text{variation due to courses}}{\text{variation due to error}}.$$

From 2.2.1 we compute:

$$p_1 = -9.34, \qquad p_4 = -4.33$$
$$p_2 = 2.67, \qquad p_5 = -1.33$$
$$p_3 = 9.33, \qquad p_6 = 3.00.$$

For the present experiment estimates of the β's if they were fixed would be

$$m_1 = \frac{1}{40}(5p_1 - p_4) = -1.059$$

$$m_2 = \frac{1}{40}(5p_2 - p_5) = .367$$

$$m_3 = \frac{1}{40}(5p_3 - p_6) = 1.091$$

2.3.1

$$m_4 = \frac{1}{40}(5p_4 - p_1) = -.308$$

$$m_5 = \frac{1}{40}(5p_5 - p_2) = -.233$$

$$m_6 = \frac{1}{40}(5p_6 - p_3) = .142$$

It should be stated that checks on the calculations of the p_i's and m_i's are possible by $\sum p_i = 0$ and $\sum m_i = 0$.

Note also that equations 2.3.1 satisfy 2.2.2 where $E(p_i)$ is replaced by p_i.

In the two methods of studying the ratio μ we will be using information which can be systematically calculated in two tables; these two tables represent a splitting up of the total sum of squares in two different ways. Computing instructions for the table entries are included in the table.

One method of studying the ratio μ is as follows:

We may decide that if the variation due to locations is less than 1/10, say, of the total variation on some particular kind of equipment

(a shoe) we won't study that item.   Thus we must decide: Is

$$\mu = \frac{\text{variation due to courses}}{\text{variation due to other causes}}$$

greater than $1/10$ or less than $1/10$?   Or, in other words, we test the hypothesis $H_0 : \mu \leq 1/10$ vs. $H_1 : \mu > 1/10$.

TABLE 2.3.1
Intra-Block Analysis

Source	Degrees of Freedom	Sum of Squares
Course (adj.)	5	$\sum_{i=1}^{6} m_i p_i = 23.1223$
Man (unadj.)	19	$\dfrac{\sum_{i=1}^{20} (\text{man total})^2}{3} - \dfrac{(\text{Grand total})^2}{60} = 116.3100$
Error	35	(subtract course and man S.S. from total) $= 261.5477$
Total	59	$\sum_{i,j} y_{ij}^2 - \dfrac{(\text{Grand total})^2}{60} = 400.9800$

TABLE 2.3.2
Inter-Block Analysis

Source	Degrees of Freedom	Sum of Squares
Course (unadj.)	5	$\dfrac{\sum_{i=1}^{6} (\text{course total})^2}{10} - \dfrac{(\text{Grand mean})^2}{60} = 53.4800$
Man (adj.)	19	subtract course and error S.S. from total $= 85.9553$
Error	35	(transfer from table I) $= 261.5477$
Total	59	(transfer from table I)

We compute that, under the assumption that μ equals 1/10, the numerator of our F statistic (the left hand side of equation 2.2.5 divided by 5) is

2.3.2
$$\left(\frac{\sum_1}{10 + 10^2\left(\frac{1}{10}\right)} + \frac{\sum_2}{\left(\frac{20}{3}\right) + \left(\frac{20}{3}\right)^2\left(\frac{1}{10}\right)}\right)\frac{1}{5} = 2.4340$$

where 10 $(= e_1)$ and 20/3 $(= e_2)$ are the characteristic roots of which we spoke in Section 2 and

$$\sum_1 = \frac{10}{10 - \frac{20}{3}}\left\{\sum_{i=1}^{6} p_i^2 - \frac{20}{3}\left(\begin{array}{l}\text{adj. sum of}\\\text{squares due}\\\text{to courses}\end{array}\right)\right\}$$

$$= 170.349$$

$$\sum_2 = \frac{\frac{20}{3}}{\frac{20}{3} - 10}\left\{\sum_{i=1}^{6} p_i^2 - 10\left(\begin{array}{l}\text{adj. sum of}\\\text{squares due}\\\text{to courses}\end{array}\right)\right\}$$

$$= 40.6$$

The mean square due to error is 7.47. The ratio of 2.3.2 to the mean square due to error is $2.434/7.47 = .326$, which is a very insignificant value for an F variate with 5 and 35 d.f. Hence, the probability of F being this large is quite high under the null hypothesis and we conclude that $\mu \leq 1/10$. We would, therefore, not study the effect of location on shoe satisfaction.

A second method of studying this ratio is to attempt to place a confidence interval on it.

We may state that in our example

$$\frac{\text{variation due to course}}{\text{variation due to error}}$$

is between 0 and .3184 with a probability of .95. The computation of this confidence interval is based on 2.2.4.

We note that

$$\Pr(\mu > \mu_{.975}) = .975$$

$$\Pr(\mu > \mu_{.025}) = .025$$

$$\Pr(\mu_{.975} < \mu < \mu_{.025}) = .975 - .025 = .95.$$

The computation of $\mu_{.025}$ according to the computing procedure of 2.2.7 is:

$$F_{.025} = .161$$

$a = (.161)(261.5477)(.143) = 6.022$

$b = 401.48674$

$\sum m_i p_i = 23.1223; \qquad \sum p_i^2 = 210.9312$

$d = -285.062 + 210.9312 = -74.1308$

$c = -17.1003$

$$\mu_{.025} = \frac{+74.1308 + \sqrt{32,957.5475}}{802.9734} = \frac{74.1308 + 181.5421}{802.9734} = .3184$$

We compute $\mu_{.975}$ in an analogous manner:

$$F_{.975} = 2.93$$

$a = (2.93)(261.5477)(.143) = 109.586$

$b = 7306.09862$

$\sum p_i^2 = 210.9312; \qquad \sum m_i p_i = 23.1223$

$d = 1441.350 + 210.9312 = 1652.2812$

$c = 86.4637$

$$\mu_{.975} = \frac{-1652.2812 + \sqrt{203,183.8939}}{14,612.1972} = \frac{-1652.2812 + 450.7593}{14,612.1972}$$

$$= -.0822$$

Then $P[\mu_{.975} < \mu < \mu_{.025}] = .95$; however, we know that μ is a ratio of squares and hence is positive. We may thus substitute 0 for μ_2 and find

$$P[0 < \mu < .318] \geq .95.$$

3. THE GENERAL CASE

3.1 *The Incomplete Block Variance Components Model.*

We now leave our illustration and generalize the computing rules so that the method may be applied to a general class of designs.

We again state our assumptions. We consider y_{ij} ($i = 1, \cdots, v$; $j = 1, \cdots, b$) to be the "yield" from the ith "treatment" and jth "block" of a statistical experiment using an incomplete block design. The reason for the quotes above is to remind the reader that these

terms may refer to applications which are not at all agricultural in nature. We further assume that the y_{ij}'s are independent and normally distributed random variables for given block effects $\beta_1, \cdots, \beta_b$ and that if the ith treatment appears in the jth block, then

$$y_{ij} = \alpha + \tau_i + \beta_j + \epsilon_{ij}$$

and the variance of y_{ij} is σ^2. In addition we assume the β's are independent and identically normally distributed with mean 0 and variance σ_1^2. Note that if the β's were unknown parameters instead of random variables that we would have the general incomplete block model with fixed effects which appears in analysis of variance (see for example Bose (1)). The total number of observations will be denoted by N, the number of treatments in each block by k, and the number of times a treatment is replicated by r. Only designs for which k and r are the same for all blocks and treatments respectively will be considered. We will denote the jth block total by B_j and the ith treatment total by T_i. Then in order to average out the treatment effects we consider the adjusted block totals

$$p_j = \begin{bmatrix} j\text{th} \\ \text{block} \\ \text{total} \end{bmatrix} - \frac{1}{k} \begin{pmatrix} \text{sum of all treatment} \\ \text{totals for treatments} \\ \text{occurring in } j\text{th block} \end{pmatrix} : j = 1, \cdots, b$$

The expectations of the p_j's are then 0 since the block effects are assumed to be random.

3.2 Linked Block Designs.

An incomplete block design has been defined to be a linked block design if

	(i)	Each block has the same number of treatments k,
3.2.1	(ii)	Each treatment occurs in r blocks,
	(iii)	Any two blocks have the same number of treatments λ^* in common.

These designs were used by Youden (6), and are duals of the well known balanced incomplete block designs.

We define

3.2.3 $$e = [k(r-1) - \lambda^*]/r$$

The illustration of Section 1 is not a linked block design but is partially linked (to be discussed in 3.3); nevertheless in a manner analogous to the illustration of 2, we may show that if F_α is the value of an F variate which has ordinate α and degrees of freedom $b - 1$ and $N - v - b + 1$,

then

3.2.4 $$\mu_\alpha = \frac{1}{e^2 F_\alpha} \cdot \frac{N - v - b + 1}{b - 1} \cdot \frac{\sum p_i^2}{\text{S.S.E.}} - \frac{1}{e}$$

is a lower bound for μ with probability α. S.S.E. is, of course, the sum of squares for error.

We may systematize the computation of $\sum p_i^2$ and S.S.E. in the following table:

TABLE 3.2.1

Source of Variation	d.f.	S.S.
Blocks eliminating treatments	$b - 1$	$\dfrac{r}{b\lambda^*} \sum p_i^2$
Treatments ignoring blocks	$v - 1$	$1/r \sum T_i^2 - (\sum y_i)^2 / N$
Error	$N - b - v + 1$	S.S.E. (by subtraction)
Total	$N - 1$	$\sum_i y_i^2 - (\sum y_i)^2 / N$

The reader may recognize this as being similar to the analysis of variance table for balanced incomplete blocks. T_i is the total for the ith treatment. We may now state the following:

Rules for linked block designs.

step i. Compute e from 3.2.3

step ii. *Confidence Intervals*

   a. compute μ_{α_1} and μ_{α_2} from 3.2.4 and table 3.2.1
   b. If $\alpha_1 < \alpha_2$, then $\mu_{\alpha_2} < \mu < \mu_{\alpha_1}$ is a confidence interval for μ with confidence coefficient $\alpha_2 - \alpha_1$. $\mu_{\alpha_2} < \mu$ is a confidence region for μ with confidence coefficient α_2 ; and $\mu < \mu_{\alpha_1}$ is a confidence region for μ with confidence coefficient $1 - \alpha_1$.

step ii'. *Size α test of $\mu \leq \mu_0$ vs $\mu > \mu_0$*

   accept $\mu \leq \mu_0$ if $F(\mu_0) = \dfrac{N - b - v + 1}{b - 1} \cdot \dfrac{1}{e + e^2 \mu_0} \cdot \dfrac{\sum p_i^2}{\text{S.S.E.}}$

   is less than F_α and accept $\mu > \mu_0$ otherwise.

3.3 *Partially Linked Designs.*

The illustration of Section 2 is a partially linked design and we here generalize the results of that section.

The dual of a Partially Balanced Incomplete Block Design is obtained by interchanging the roles of the treatments and blocks. In analogy with linked block designs we may call these dual designs partially linked designs. The following conditions are satisfied:

(i) The experimental material is divided into b blocks of k units each, different treatments being applied to the units in the same block.

(ii) There are v treatments, each of which occurs in r blocks.

(iii) There can be established a relation of association between any two blocks satisfying the following requirements:

a) Two blocks are either 1st, 2nd, $\cdots$, or m*th associates.

b) Each block has exactly n_i^*, ith associates ($i = 1, 2, \cdots, m^*$).

c) Given any two blocks which are ith associates, the number of blocks common to the jth associates of the first, and the kth associates of the second is $p_{jk}'^*$ and is independent of the pair of blocks with which we start. Also $p_{jk}^{i*} = p_{kj}^{i*}$.

(iv) Two blocks which are ith associates contain exactly λ_i^* common treatments.

If the dual of the design we are working with is tabulated in Bose, Clatworthy and Shrikhande (2) then H^* and Δ^* are the H and Δ tabulated there for this dual design. In the present framework, the key result for partially linked designs is that if F_α is the value of an F variate which has ordinate α and degrees of freedom $b - 1$ and $N - v - b + 1$; and if

$$a_\alpha = F_\alpha \begin{Bmatrix} \text{sum of} \\ \text{squares} \\ \text{due to} \\ \text{error} \end{Bmatrix} \frac{b - 1}{N - b - v + 1}$$

$$b_\alpha = a_s \Delta^*$$

3.3.1

$$c_\alpha = a_s - \begin{Bmatrix} \text{adjusted} \\ \text{sum of} \\ \text{squares for} \\ \text{blocks} \end{Bmatrix}$$

$$d_\alpha = c_\alpha H^* + \sum p_i^2$$

then

3.3.2
$$\mu_\alpha = \frac{-d_\alpha + \sqrt{d_\alpha^2 - 4b_\alpha c_\alpha}}{2b_\alpha}$$

is a lower bound for μ with probability α.

We systematize the computation of the adjusted sum of squares for blocks, $\sum p_i^2$, and S.S.E. in the following two tables.

TABLE 3.3.1

Source	d.f.	Sum of Squares
Blocks (adj.)	$b - 1$	$\sum_{j=1}^{b} m_j p_j$
Treatments (unadj.)	$v - 1$	$\sum_{i=1}^{v} \frac{T_i^2}{r} - \frac{G^2}{N}$
Error	$N - b - v + 1$	S_e^2 (by subtraction)
Total	$N - 1$	$\sum_{i,j} y_{ij}^2 - \frac{G^2}{N}$

TABLE 3.3.2

Source	d.f.	Sum of Squares
Blocks (unadj.)	$b - 1$	$\frac{1}{k} \sum_{j=1}^{b} B_j^2 - \frac{G^2}{N}$
Treatments (adj.)	$v - 1$	S_t^2 (by subtraction)
Error	$N - v - b + 1$	S_e^2 (by transfer from table 3.3.1)
Total	$N - 1$	$\sum y_{ij}^2 - \frac{G^2}{N}$

Here m_1, m_2, $\cdots$, m_b are a solution to the equations

$$d_{11}m_1 + d_{12}m_2 + \cdots + d_{1b}m_b = p_1$$
$$d_{12}m_1 + d_{22}m_2 + \cdots + d_{2b}m_b = p_2$$

$$\qquad \cdot \qquad\quad \cdot \qquad\qquad \cdot \qquad \cdot$$
$$\qquad \cdot \qquad\quad \cdot \qquad\qquad \cdot \qquad \cdot$$
$$\qquad \cdot \qquad\quad \cdot \qquad\qquad \cdot \qquad \cdot$$

$$d_{1b}m_1 + d_{2b}m_2 + \cdots + d_{bb}m_b = p_b$$

Hence $m_1 - m_2$ would estimate $\beta_1 - \beta_2$ if the β's were fixed instead of random. G is the grand total of the y_{ij}'s.

We summarize:

rules for partially linked designs $(m^* = 2)$

i. *Two sided confidence interval*
   a. If possible, look up the dual of the design you are using in the B.C.S. catalog (2). Set H and Δ found there equal to H^* and Δ^* respectively.
   b. Compute S.S.E. and $\sum m_i p_i$ from tables 3.3.1 and 3.3.2.
   c. Compute a_{α_1}, b_{α_1}, d_{α_1}, c_{α_1} and μ_{α_1} from 3.3.1 and 3.3.2 in turn. Do the same for μ_{α_2}.
   d. If $\alpha_1 < \alpha_2$, then $\mu_{\alpha_2} < \mu < \mu_{\alpha_1}$ is a confidence interval for μ with confidence coefficient $\alpha_2 - \alpha_1$.
ii. *One sided confidence interval*
   a. ⎫
   b. ⎬ same as in *i*.
   c. ⎭
   d. $\mu_{\alpha_2} < \mu$ is a confidence region for μ with confidence coefficient α_2; and $\mu < \mu_{\alpha_1}$ is a confidence region for μ with confidence coefficient $1 - \alpha_1$.
iii. *Test of* $\mu \leq \mu_0$ *vs* $\mu > \mu_0$
   accept $\mu \leq \mu_0$ if

$$\frac{N - b - v + 1}{b - 1} \cdot \frac{1}{\text{S.S.E.}} \cdot \frac{(\sum m_i p_i)[1 + \mu_0 H^*] - \mu_0 \sum p_i^2}{1 + \mu_0 H^* + \mu_0^2 \Delta^*}$$

is less than F_α and accept $\mu > \mu_0$ otherwise.

3.4 A note of warning.

It is to be emphasized that the above rules apply only to linked block and partially linked designs $(m^* = 2)$ whose duals are listed in

the B.C.S. catalog. This latter group, however, includes essentially all known partially linked designs with $m^* = 2$ except the duals of lattices. Our procedure will then be as follows: In setting up the experiment pick a design which is (i) balanced or partially balanced $(m = 2)$ and (ii) linked or partially linked $(m^* = 2)$; then, if the experiment is carried out intact we may use the rules of this section to analyze the variance of the block effects and the well known procedures of inter-block analysis to analyze the treatment effects.

If we were to use an arbitrary design we would, in general, find it difficult to identify the nature of the dual and hence would not be able to apply the rules of this section.

<div align="center">4. SECOND EXAMPLE</div>

4.1 *Introduction*

It has long been recognized that designing experiments with a view to analyzing the "treatment" effects simplifies the analysis greatly; it isn't surprising then, that if we want to analyze the "block" variability that such an analysis is greatly simplified by designing the experiment with this in mind.

Now, partially balancing a design is a method of designing an experiment to simplify the treatment analysis and it appears from the preceding sections that partially linking is a method of simplifying the block error analysis. Fortunately designs may be both partially balanced and partially linked. One such "nice" design from this point of view will be analyzed to illustrate these general remarks.

4.2 *A Linked Block Design*

As an illustrative example we will give the analysis of an experiment on the yields of 35 varieties of oats by Dr. G. K. Middleton. This is Design 2, Table III A of Bose and Shimamoto (3). The treatment analysis of this experiment may be found in Bose, Clatworthy, and Shrikhande (2). This is a particularly interesting design from the present point of view, because it is both partially balanced and a linked block design. The dual of the Bose and Shimamoto Design 2, Table III A is the balanced incomplete block design with parameters $u = 15$, $b = 35$, $r = 7$, $k = 3$ and $\lambda = 1$, hence we follow the rules of 3.2 using:

$$b = 15 \qquad\qquad u = 35$$

$$k = 7 \qquad\qquad r = 3$$

$$\lambda^* = 1$$

TABLE 4.2.1

Treat. \ Blocks	1	2	3	4	5	6	7	8
1	442							
2		442					556	
3				407				
4	502		498				580	
5		456				504		
6						515		406
7				526				
8		452			443			
9				440		526		
10			320					
11		482						
12					280			365
13							434	
14					417			
15	314							260
16							491	
17	428							
18	512	427		413				
19		436	378					
20						314	366	
21			272	306				315
22			413			286		
23					265	271		
24								
25				212				
26		431			326			
27								
28							284	285
29								365
30	404				328			
31		411						380
32								
33				385	286		370	
34			334					
35	389					455		
Block Total = B_i	2991	3085	2667	2689	2345	2871	3081	2376
$B_i - p_i$	2920	3077	2645	2830	2568	2910	2909	2468
p_i	71	8	22	−141	−223	−39	172	−92

TABLE 4.2.1 (*Cont.*)

9	10	11	12	13	14	15	Treat. Total	Treat. Means	Treat.
	400					438	1280	427	1
		347					1345	488	2
	452			498			1357	452	3
							1580	527	4
				576			1536	512	5
		592					1513	504	6
			642			522	1690	563	7
		496					1391	464	8
					433		1399	466	9
325						345	990	330	10
480			532				1494	498	11
	364						1009	336	12
			376		431		1241	414	13
328				403			1148	383	14
					244		818	273	15
				634		488	1613	538	16
		431		486			1345	448	17
							1352	451	18
	345						1159	386	19
	430						1110	370	20
							893	298	21
			386				1085	362	22
						260	796	265	23
		275			276	312	863	288	24
292		254					758	253	25
					472		1229	410	26
325	322				352		999	333	27
226							795	265	28
			476	418			1259	420	29
			356				1088	363	30
						326	1117	372	31
	425	510	407				1342	447	32
							1041	347	33
				265	235		834	278	34
450							1294	431	35
2426	2738	2905	3175	3280	2443	2691	41,763	13,922	
2493	2751	2852	3067	3031	2462	2783	41,766		
−67	−13	53	108	249	−19	−92	−3		

4.2.1 is a table in which the yields of oats are cross classified according to what treatment they represent and what block they were planted in. The column sums are then the block totals, B_i, and the row sums are the treatment totals, T_i. A check is furnished by summing the block totals and the treatment totals, since both sums should equal the grand total, G. An additional column to the right may be used to calculate the treatment means T_i/k, and this may again be checked by summing and comparing with the grand mean since $\sum T_i/k = G/k$. We may also use this table to compute the p_i's. We first compute $B_i - p_i$ by summing the means of the treatments appearing in the ith column. As an example

$$B_1 - p_1 = 427 + 527 + 273 + 448 + 451 + 363 + 431$$

$$= 2920$$

p_i is then calculated by subtracting $B_i - p_i$ from B_i. $B_1 = 2991 - 2920 = 71$. Checks are furnished at each of the last two stages by resorting to the identities $\sum_i (B_i - p_i) = G$ and $\sum_i p_i = 0$, respectively.

In the present numerical example Table 3.2.1 becomes

TABLE 4.2.2

Source of Variation	d.f.	S.S.
Blocks eliminating treatments	14	$\dfrac{3}{15.1}$ 205,165
Treatments ignoring blocks	34	753,254
Error	56	141,155
Total	104	935,442

From 3.2.3 we see that

$$e = \frac{1}{r}\,[k(r-1) - \lambda^*]$$

$$= \frac{1}{3}\,[(7)(2) - 1]$$

$$= \frac{13}{3}\,.$$

Let $\alpha_1 = .05$ and $\alpha_2 = .95$, then $F_{\alpha_2} = 1.75$ and $F_{\alpha_1} = 1/2.11 = .475$. Now from 3.2.4 we see that

$$\mu_s = \frac{1}{\left(\dfrac{13}{3}\right)^2 F_s} \cdot \frac{104}{14} \cdot \frac{205{,}165}{141{,}155} - \frac{1}{\dfrac{13}{3}} \; ; \qquad s = \alpha_1 \, , \alpha_2$$

$$\mu_{\alpha_1} = \quad .097$$

$$\mu_{\alpha_2} = \; 1.884.$$

And finally (.097, 1.884) is a 90% confidence interval for μ, (0, 1.884) and (.097, ∞) are 95% confidence regions for μ.

If we wish to perform the test of step ii$'$ with $\mu_0 = 3$, say, and significance level .05; then we compute $F(3)$

$$F(3) = \frac{104}{14} \frac{1}{\dfrac{13}{3} + \left(\dfrac{13}{3}\right)^2 3} \frac{205{,}165}{141{,}155}$$

$$= .178.$$

Now since $F(3) = .178 < F_{.05} = .475$, we accept the hypothesis that $\mu \leq 3$. It is easily seen that if μ_0 has any value greater than 1.884 then we will accept the hypothesis $\mu \leq \mu_0$.

REFERENCES

1. R. C. Bose, "Least Square Aspects of Analysis of Variance," Institute of Statistics, University of North Carolina.
2. R. C. Bose, W. H. Clatworthy and S. S. Shrikhande, "Tables of Partially Balanced Designs with Two Associative Classes," Institute of Statistics, University of North Carolina.
3. R. C. Bose and T. Shimamoto, "Classification and Analysis of Partially Balanced Incomplete Block Designs with Two Associate Classes," *Journal of the American Statistical Association*, Vol. 47, No. 258.
4. W. A. Thompson, Jr., "Statistical Analysis of Military Risks Resulting from Extreme Environmental Conditions," Report No. 3, 1954.
5. Abraham Wald, "A Note on Regression Analysis," *Annals of Mathematical Statistics*, Vol. 18, p. 586.
6. W. J. Youden, "Linked Blocks: A New Class of Incomplete Block Designs," *Biometrics*, 7 (1951), 124 (abstract).

A NOTE ON DESIGN AND ANALYSIS OF SOIL INSECTICIDE EXPERIMENTS

P. SPRENT

University of Tasmania

In a paper in this journal, van der Reyden [2] proposes a design in which control plots, one adjacent to each treated plot, are used to obtain adjustments for uneven distribution of insects in the soil. He discusses the application of his method to a soil insecticide experiment in rectangular lattice design. It unfortunately proved impossible to obtain the original data from van der Reyden's experiment, so that the present writer was unable to clear up certain points of ambiguity in the necessarily brief description, or to try different alternative methods of analysis. In this note attention is drawn to certain aspects of the method which do not seem satisfactory.

Van der Reyden's method of analysis appears to consist of analysing separately the results for the treated and control plots, obtaining the estimated treatment effects and residuals for each plot. Two corrections are then applied to the observed value for each treatment plot, viz:

(i) the value of the "treatment" constant of the corresponding control plot is subtracted from the treatment constant for the the treated plot,

(ii) the residual of the treated plot is replaced by the mean of the residuals of the treated plot and the corresponding control plot.

"Corrected" values for each treated plot are then reconstructed from the mean, replicate, block and treatment constants and residuals, modified as above. The results are subjected to an ordinary analysis of variance.

With regard to correction (i) there is some ambiguity in van der Reyden's paper. It is implied on p. 293 that the "treatment" constant for the control plot is added rather than subtracted. There is further evidence on this point, to be discussed later, in the numerical example. It suffices to point out here that if we regard these control "treatment" effects as estimates of the infestation on the corresponding treated plots, then what we measure on the treated plots as treatment is a "treatment + infestation" effect. The obvious correction is to subtract our estimate of infestation.

It should be noted that correction (i) corresponds to the correction used in the ordinary analysis of covariance when the regression coefficient, b, is taken as unity. In this case the residuals for the corrected values are $\epsilon_u - \epsilon_w$, and the corresponding error variance is

$$V(\epsilon_u) + V(\epsilon_w) - 2 \text{ cov } (\epsilon_u , \epsilon_w) \tag{1}$$

where ϵ_u and ϵ_w are the residuals for the treated and control plot respectively.

It should be noted that the appropriate residuals and the corresponding error variances are determined by our procedure of estimation. Had we used some other procedure (such as the orthodox covariance technique) we would have been led to different, but appropriate, values of these residuals.

With van der Reyden's correction (ii) the residuals will be $\frac{1}{2}(\epsilon_u + \epsilon_w)$, and the corresponding error variance given by the analysis of variance of the corrected values will be

$$\tfrac{1}{4}\{V(\epsilon_u) + V(\epsilon_w) + 2 \text{ cov } (\epsilon_u , \epsilon_w)\} \tag{2}$$

Obviously expressions (1) and (2) will not in general be equivalent, and consequently the analysis of variance of the corrected values does not give a correct estimate of the error to which these values are subject.

Inspection of van der Reyden's Table II lends support to the view that his value for the error variance of the corrected values is a serious underestimate. If cov $(\epsilon_u , \epsilon_w)$ is zero the correct error variance would be four times that given by van der Reyden's method. In this case van der Reyden's value would be one quarter of the sum of the control and treated error variances, namely 13.0. The value actually obtained by van der Reyden, 9.4, thus suggests that cov $(\epsilon_u , \epsilon_w)$ has a small negative value.

A number of effects and interactions have been judged highly significant for the corrected data. There seems little scientific or agricultural reason why some of these should be so. For instance, the interactions RT and $R'T$ are each highly significant, whereas neither R nor R' reach significance at the 5% level. Yates [3] has pointed out that unexpected significance of interactions may be a warning of faulty analytical procedure.

Further, inspection of van der Reyden's Table I gives little indication of the trends one would expect if the treatment differences really were significant. For instance, there are no very noticeable trends associated with increasing concentration of chemicals, or with time of application.

As already mentioned, there is some evidence that correction (i) has been incorrectly applied to the numerical example. An inspection of total stand losses for two replicates as given by van der Reyden indicates that the correction has been added rather than subtracted. It will be noted from his Table I that when stand losses are high on controls the correction is nearly always positive, the opposite occuring when control stand losses are low. Indeed, if the entries are arranged in descending order of stand losses for control plots, in the first 21 cases the correction results in an increased figure in all but two cases. For the remaining 21 cases the correction is zero or negative in all but two cases. The net effect is to give a greater spread of stand losses, resulting in an increase of both treatment and total sums of squares, as is evident from van der Reyden's Table II.

A further point of criticism of van der Reyden's procedure arises from his statement that if "treatment" effects on the control plots do not prove to be significant, no further attention need be given to the data from them. This recommendation seems to illustrate a misconception.

The non-random distribution of insects over the experimental area only represents in an aggravated form the same problem as arises from the non-random distribution of soil fertility in a fertilizer experiment or variety trial. The use of randomisation, as explained by Fisher [1] is designed to overcome this difficulty. If randomisation has been correctly carried out the straightforward analysis of the treated plots will be *valid*, even though the *precision* of the experiment may not be high. Furthermore, if correctly randomised, the "treatment" effects in an analysis of the control plots will only prove significant in 5% of all trials on the average, if the 5% significance level is chosen, unless there has been some carry over of treatment effects from neighboring plots.

It may well be possible to improve the precision of the experiment by using the results from the control plots. In general, the appropriate statistical technique is the analysis of covariance. Van der Reyden's objection to this, that "the more effective a treatment, the less the correlation between treated and control plots" would appear to have relevance only in the case of highly effective treatments which reduce stand losses to zero. What matters is the correlation between the residuals on the treated and control plots, and except in the case just mentioned this may well be high regardless of treatment effects.

Whereas van der Reyden recommends using control results only when "treatment" affects are significant on control plots, in this case the very fact that they are significant should be taken as a warning

to proceed with caution, in case, as already suggested, the treatments have affected control plot observations by a carry over effect. Before control plot observations, or other sources, are used to provide supplementary information the experimenter must satisfy himself that this information does not itself reflect treatment effects. For instance, in field trials, counts of the numbers of seeds germinating, or of numbers of plants maturing, should not be used as supplementary information if these numbers are themselves influenced by the treatments under investigation. On the other hand, in a field trial it would be perfectly safe to use as supplementary information the yields of a uniformity trial conducted on the same land in the year previous to that in which the treatments under test were applied, since the treatments cannot have affected these yields.

In conclusion, it may be remarked that the irregular distribution of insects in experiments such as that described by van der Reyden raises important problems of technique if useful information is to be obtained. This writer does not know if it was practicable in this case to use insect counts on soil samples to obtain supplementary information. This technique has been widely used.

Acknowledgements

REFERENCES

1. Fisher, R. A. *The Design of Experiments*. Oliver & Boyd, 6th Edn., 1951.
2. van der Reyden, D. Design and Analysis of Soil Insecticide Field Experiments. *Biometrics*, *10*, 291, 1954.
3. Yates, F. The Principles of Orthogonality and Confounding in Replicated Experiments. *J. Agri. Sci.*, *23*, 108, 1933.

COVARIANCE ANALYSIS AS AN ALTERNATIVE TO STRATIFICATION IN THE CONTROL OF GRADIENTS

ANNE D. OUTHWAITE AND A. RUTHERFORD*

Department of Statistics, University of Aberdeen

Introduction

In a recent paper, Federer and Schlottfeldt (1) illustrated the use of covariance to control gradients in an experiment as a substitute for deliberate stratification in the design. For this purpose, they took account of linear and quadratic trends. Since there was no obvious reason for stopping at this stage, we have examined the effect of including all terms up to the sixth degree.

The Covariance Analysis

Federer and Schlottfeldt discussed measurements of the heights of tobacco plants in an experiment on seven treatments arranged in eight randomized blocks. A fertility gradient within the blocks was suspected and they therefore calculated a quadratic covariance analysis on a measure of distance in this direction. For the study of a regression trend of higher degree, the computations are simplified by using standard orthogonal polynomial values from Fisher and Yates's Statistical Tables (2) based upon distance from the centre of the experiment, this modification involving no difference in principle. Table I reproduces the yields from (1) and also the covariates, x_1 to x_6 for the corresponding orthogonal polynomials.

The analysis of squares and products up to the third degree is shown in Table II, which includes the quantities required for subsequent covariance adjustments and agrees with Tables III and IV of (1) with respect to x_1, x_2 and y.

To estimate the regression coefficients in the cubic analysis, the following set of equations must be solved.

$$214.750b_1 + 9.500b_2 - 5.375b_3 = 4,559.7$$

$$9.500b_1 + 585.250b_2 - 7.625b_3 = 17,906.6$$

$$-5.375b_1 - 7.625b_2 + 38.000b_3 = -3,238.6$$

*On study leave from Ministry of Agriculture, N. Ireland.

TABLE I

Plant Height, First Measurement (Total for 20 Plants in cms)

Covariate						Replicate Number								Totals
x_1	x_2	x_3	x_4	x_5	x_6	1	2	3	4	5	6	7	8	
−3	5	−1	3	−1	−1	F 1299.2	B 1369.2	A 1169.5	F 1219.1	F 1120.0	.G 1031.5	B 1076.4	D 1099.6	9384.5
−2	0	1	−7	4	6	G 875.9	E 844.2	E 975.8	A 971.7	G 827.0	B 846.5	A 917.9	E 947.4	7206.4
−1	−3	1	1	−5	−15	D 960.7	F 968.7	C 873.4	G 607.6	D 671.9	D 667.8	E 627.6	B 787.1	6164.8
0	−4	0	6	0	20	C 1004.0	G 975.5	G 797.8	D 1000.0	C 972.2	C 853.6	F 776.4	A 898.3	7277.8
1	−3	−1	1	5	−15	A 1173.2	C 1322.4	B 1069.7	B 1343.3	A 1083.7	A 1087.1	C 960.4	C 1174.9	9214.7
2	0	−1	−7	−4	6	B 1031.9	A 1172.6	F 1093.3	E 999.4	B 1146.9	E 990.2	G 852.4	F 1003.3	8290.0
3	5	1	3	1	−1	E 1421.1	D 1418.9	D 1169.6	C 1181.3	E 993.8	F 1021.9	D 1006.2	G 947.6	9160.4
0	0	0	0	0	0	7766.0	8071.5	7149.1	7322.4	6815.5	6498.6	6217.3	6858.2	56698.6

TABLE II .

Cubic Covariance Analysis

Source	D.F.	$\sum x_1^2$	$\sum x_2^2$	$\sum x_3^2$	$\sum x_1x_2$	$\sum x_1x_3$	$\sum x_2x_3$	$\sum x_{1y}$	$\sum x_{2y}$	$\sum x_{3y}$	$\sum y^2$
Total	55	224.00	672.00	48.00	0.000	0.000	0.000	4544.8	17414.8	−4357.6	1,931,777
Replicates	7	0.00	0.00	0.00	0.000	0.000	0.000	0.0	0.0	0.0	388,314
Treatments	6	9.25	86.75	10.00	−9.500	5.375	7.625	−14.9	−431.8	−1119.0	273,875
Error	42	214.75	585.25	38.00	9.500	−5.375	−7.625	4559.7	17906.6	−3238.6	1,269,588
$E + T$	48	224.00	672.00	48.00	0	0	0	4544.8	17474.8	−4357.6	1,543,463

TABLE III
Regression Analysis

Degree of Polynomial	D.F.	Sum of Squares removed by Regression Analysis					
		First Degree	Second Degree	Third Degree	Fourth Degree	Fifth Degree	Sixth Degree
First	1	96816					
Second	1		535991				
Third	1			222767			
Fourth	1				17088		
Fifth	1					130116	
Sixth	1						17385
D.F. for Error S.S.	42	41	40	39	38	37	36
Error Sum of Squares	1269588	1172772	636781	414014	396926	266810	249425
Error Mean Square	30228	28604	15920	10616	10445	7211	6928
F		3.38	33.67	20.99	1.63	18.04	2.51

As the variances of the adjusted means were required, it was simplest to invert the matrix of the coefficients giving:—

$$(c_{ij}) = \begin{bmatrix} 0.0046758 & -0.0000675 & 0.0006478 \\ -0.0000675 & 0.0017141 & 0.0003344 \\ 0.0006478 & 0.0003344 & 0.0264745 \end{bmatrix}$$

and so:—

$$b_1 = 18.01396$$
$$b_2 = 29.30342$$
$$b_3 = -76.79897$$

The same process was followed for the regression coefficients in the higher degree covariance analyses. Table III shows the amount by which the error sum of squares was reduced by successive steps in the analysis up to the sixth degree term. The first entry in each column is obtained by subtracting the error sum of squares for that term from the previous error sum of squares. The error sum of squares for the quadratic differs from that in (1) because of an arithmetical error in the original paper. Tests of significance made by comparing the square for each term of the regression with the corresponding error mean square are open to criticism in that successive tests are not independent, but they strongly indicate that a covariance adjustment ought to include the third and fifth degree components though the fourth and sixth are of little importance for this set of data.

The Adjusted Means

As orthogonal polynomials were used, each $\bar{x}_i$ was zero. Hence the formulae for the adjusted treatment means $\bar{Y}_i'$, (where the $\bar{Y}_i$ are the unadjusted treatment means) are

$$\bar{Y}_i' = \bar{Y}_i - \sum_{j=1}^{n} b_j \bar{x}_{ji}$$

for the nth degree covariance where the b_j are the regression coefficients calculated for each covariance analysis, as given in Table IV.

From these, the means in Table V are obtained; these have been given only to one decimal place, as there appears to be little object in going further, when the standard errors are over 30 even after the sixth degree covariance adjustment.

TABLE IV

Regression Coefficients

Degree of Covariance Analysis	b_1	b_2	b_3	b_4	b_5	b_6
First	21.22934					
Second	19.89326	30.27350				
Third	18.01396	29.30342	−76.79897			
Fourth	18.40304	28.94409	−76.48360	4.21134		
Fifth	18.65597	26.84553	−80.88729	3.81832	15.34852	
Sixth	18.75972	26.80314	−81.34208	3.94426	15.32957	−1.58610

Variance of Adjusted Means

The general formula for the variance of the difference between two means adjusted for linear regression is

$$V(\bar{Y}_1' - \bar{Y}_2') = s^2 \left(\frac{2}{r} + \frac{(\bar{x}_{1i} - \bar{x}_{2i})^2}{A} \right)$$

where s^2 is the residual error mean square for y after the removal of the regression component, r is the number of replicates, and A is the error sum of squares for x. This variance depends on the pair of treatments compared; Finney (3) pointed out that, if treatment differences in x are fairly small, this inconvenience could be avoided by averaging the second term over all possible pairs. He showed this to be equivalent to taking the variance of any one adjusted mean as

$$V(\bar{Y}_i') = \frac{s^2}{r} \left(1 + \frac{d}{A(t-1)} \right)$$

where d is the sum of squares for "treatments" for x, and t is the number of treatments. By generalising this the following formula is found for nth degree polynomial:

$$V(\bar{Y}_i') = \frac{s^2}{r} \left(1 + \frac{1}{(t-1)} \sum_{i,j=1}^{n} c_{ij} d_{ij} \right)$$

where the d_{ij} is the "treatment" sum of products of x_i and x_j, and the c_{ij} are the elements of the inverse matrix of the error sums of squares and sums of products.

Gains in Information

Federer and Schlottfeldt (1) compared the information obtained from alternative covariance analyses essentially in terms of s^2, without

TABLE V

Unadjusted Treatment Totals and Adjusted Treatment Means of Plant Height, for Second, Third and Fifth Degree

Treatments	Treatment Totals						Means for 20 Plants (cms)		
	x_{1i}	x_{2i}	x_{3i}	x_{4i}	x_{5i}	Y_i	Second Degree Adjustment	Third Degree Adjustment	Fifth Degree Adjustment
A	−2	−8	−3	−9	18	8474.0	1094.5	1064.3	1030.2
B	−3	1	−4	−12	−1	8671.0	1087.6	1048.6	1054.7
C	5	−19	−1	25	11	8342.2	1102.2	1091.5	1051.7
D	3	7	5	21	−13	7994.7	965.4	1014.9	1034.3
E	3	7	4	−28	1	7799.5	941.0	980.9	996.3
F	−3	13	−3	5	−15	8501.9	1021.0	993.1	1022.2
G	−3	−1	2	−2	−1	6915.3	875.7	894.0	897.9
Standard Errors							±45.3	±37.8	±32.4

allowance for the errors of estimation of regression coefficients. Table VI illustrates the consistent underestimation of variances to which this leads, the omitted terms being always positive. We consider that the most easily interpreted measure of gain in information is given by a comparison of variances of treatment means with different covariance adjustments, always making use of the formula of the last section. Table VII shows these relative efficiencies and also, to compare with the original paper, the units of information assessed on a basis of a standard error equal to 5% of the general mean.

TABLE VI

Variance of a Treatment Adjusted for Covariance
With and Without Corrections

Degree of Covariance Adjustment	Without Corrections	With Corrections
Unadjusted	3778.5	3778.5
First	3575.5	3601.2
Second	1990.0	2053.2
Third	1327.0	1431.0
Fourth	1305.6	1470.3
Fifth	901.4	1052.4
Sixth	866.0	1021.4

TABLE VII

Comparison With Unadjusted Variance and Gains in Information

Degree of Covariance Adjustment	Comparisons With	
	Unadjusted Variance of Y	A Precision Equal to 5% of Mean
Unadjusted	1.00	0.65
First	1.05	0.68
Second	1.84	1.19
Third	2.64	1.71
Fourth	2.57	1.66
Fifth	3.59	2.31
Sixth	3.70	2.38

Comparison with Latin Square

If an experimental area were suspected of having fertility trends in two directions, the natural design to adopt would be a Latin square or

some modification of this. It is therefore natural to enquire how a 7 × 7 Latin square would have compared with the eight randomized blocks actually used in this experiment. Our analysis may be regarded as an attempt to obtain some of the advantages of a Latin square from an experiment in randomized blocks. Indeed, if a Latin square design were first analysed by calculating sums of squares for treatments and for columns but not for rows, and then a multiple covariance were used to eliminate polynomial trends of the highest possible degree between rows, the outcome would be exactly the same as that of the ordinary Latin square analysis; the calculations described above would become trivial. The orthogonality property of the Latin square makes valid the much more convenient procedure of direct calculation of a sum of squares between rows.

The residual error mean square obtained in our analysis, after eliminating all polynomial trends up to the sixth degree, estimates the mean square that would have been obtained if a 7 × 7 Latin square had been used on the same area with the same size of plot (of course omitting seven of the plots, or one block of the actual experiment). Consequently the error variance of a treatment mean for such a Latin square experiment may be estimated to be 989.7, which may be compared with the final figure obtained for the average variance of a treatment mean in the randomized block design 1021.4. Despite the additional replication, the randomized block design seems to compare the treatments with slightly less precision (at least in respect of this one measurement) than a Latin square design would have done. The corrections to the variance in the randomized block experiment, arising from sampling errors in the estimation of regression coefficients, more than counterbalance the gain of one replicate. The randomized block design has the further advantage of estimating the error with 36 degrees of freedom instead of 30 from the Latin square, but, when the number of degrees of freedom is of this order, such an increase is scarcely worth having.

On the evidence of this experiment then, nothing has been gained by having one replicate more than the Latin square would give, as the increase in precision has been lost by the destruction of orthogonality. This finding for the analysis of one measurement in one experiment does not establish any general principle; it does suggest that experimenters should be cautious in departing from a simple orthogonal design merely to achieve a slight increase in replication.

In the circumstances of this experiment, a compromise might have been effected by constructing the design as a 7 × 7 Latin square with one additional column. Cochran and Cox (4) have discussed such designs in their Chapter 13, and have shown that the design with seven

treatments and eight replicates has 98% efficiency; their measure of efficiency can be shown to be equal to the factor

$$\frac{(t-1)}{(t-1) + \sum c_{ij}d_{ij}}$$

In the design used, however, the corresponding efficiency was only 85%. Thus as compared with the 7×7 Latin square the extra replication in the Cochran and Cox design would give a net gain in precision of 12% ($0.98 \times 8/7 = 1.12$), while for the design used the corresponding figure was a net loss of 3% ($0.85 \times 8/7 = 0.97$)

Summary

Federer and Schlottfeldt (1) discussed the analysis of a randomized block design with eight blocks and seven treatments in which there was a fertility trend within the blocks. They used a covariance analysis to eliminate linear and quadratic components of this trend. We have extended this to the limit, using orthogonal polynomials to the sixth degree. We have pointed out that the variances of the adjusted means should be corrected to allow for the sampling errors of the regression coefficients, and have given a general formula for the average value of the corrections. These corrected variances were used to calculate the gains in information from the covariance analysis. Even after the full covariance adjustment, the design used turns out to be less efficient than a 7×7 Latin square with the same variance per plot, as the advantage gained by having an extra replicate was lost by the extreme non-orthogonality. For similar circumstances, Cochran and Cox (4) have suggested a design with eight replicates consisting of a 7×7 Latin square with an extra column; this design would have been a better choice, since it departs so little from orthogonality that the gain in using the extra replicate outweighs the loss due to nonorthogonality and leaves an overall gain in precision of 12% (assuming equal variances per plot).

REFERENCES

(1) Federer, W. T. & Schlottfeldt, C. S. (1954): The use of covariance to control gradients in experiments. *Biometrics*, *10*, 282–290.

(2) Fisher, R. A. & Yates, F. (1953): *Statistical Tables for Biological, Agricultural and Medical Research.* (3rd edition). London: Oliver & Boyd.

(3) Finney, D. J. (1946): Standard errors of yields adjusted for regression on an independent measurement. *Biometrics*, *2*, 53–55.

(4) Cochran, W. G. & Cox, G. M. (1950): *Experimental Designs.* New York: John Wiley & Sons.

ON THE ANALYSIS OF VARIANCE OF A TWO-WAY CLASSIFICATION WITH UNEQUAL SUB-CLASS NUMBERS

CLYDE YOUNG KRAMER

Virginia Agricultural Experiment Station
Blacksburg, Virginia

In many avenues of research it is necessary to analyse the variance of data which are classified in two ways with unequal numbers of observations falling into each sub-class of the classification. For data of this kind special methods of analysis are required because the inequality of the sub-class numbers causes lack of orthogonality among the main effects and interaction comparisons.

Table I below gives the basic notation for dealing with an analysis of a two-way classification with unequal sub-class numbers.

Several writers have dealt with the analysis of data of this form and various methods have been put forward. Some of the more prominent articles and discussions are cited below [1–13].

A simple preliminary step common to all methods is to separate the variance within sub-classes from the variance between sub-classes. Table II gives the analysis of variance for this preliminary step.

The problem of extending the analysis to the main effects and to the interaction between the main effects now arises. The $(pq - 1)$ degrees of freedom for between sub-classes can be partitioned in the usual way into $(p - 1)$ degrees of freedom for between A classes, $(q - 1)$ degrees of freedom for between B classes and $(p - 1) (q - 1)$ degrees of freedom for the interaction between the two classifications. The main difficulties arise in determining the correct sums of squares to be associated with each of these. One difficulty is that the addition theorem for sums of squares does not apply unless the sub-class numbers are proportional, and thus the interaction sum of squares cannot be computed by the usual method of differences. In fact, situations may occur where this procedure would give a negative result for the sum of squares for interaction.

Frequently we assume, from the nature of the data or from previous information or experience, that interaction is absent or if present, negligible. Making this assumption, we are interested in testing if there are any significant differences between the A-classes and between the B-classes.

TABLE I

Basic Notation

A Classes		B Classes						Total
		B_1	B_2	$\cdots$	B_j	$\cdots$	B_q	Total
A_1	No.	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1q}	$n_1.$
	Total	Y_{11}	Y_{12}	$\cdots$	Y_{1j}	$\cdots$	Y_{1q}	$Y_1.$
	Mean	y_{11}	y_{12}	$\cdots$	y_{1j}	$\cdots$	y_{1q}	$y_1.$
A_2	No.	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2q}	$n_2.$
	Total	Y_{21}	Y_{22}	$\cdots$	Y_{2j}	$\cdots$	Y_{2q}	$Y_2.$
	Mean	y_{21}	y_{22}	$\cdots$	y_{2j}	$\cdots$	y_{2q}	$y_2.$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
A_i	No.	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iq}	$n_i.$
	Total	Y_{i1}	Y_{i2}	$\cdots$	Y_{ij}	$\cdots$	Y_{iq}	$Y_i.$
	Mean	y_{i1}	y_{i2}	$\cdots$	y_{ij}	$\cdots$	y_{iq}	$y_i.$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
A_p	No.	n_{p1}	n_{p2}	$\cdots$	n_{pj}	$\cdots$	n_{pq}	$n_p.$
	Total	Y_{p1}	Y_{p2}	$\cdots$	Y_{pj}	$\cdots$	Y_{pq}	$Y_p.$
	Mean	y_{p1}	y_{p2}	$\cdots$	y_{pj}	$\cdots$	y_{pq}	$y_p.$
Total	No.	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.j}$	$\cdots$	$n_{.q}$	$n_{..}$
	Total	$Y_{.1}$	$Y_{.2}$	$\cdots$	$Y_{.j}$	$\cdots$	$Y_{.q}$	$Y_{..}$
	Mean	$y_{.1}$	$y_{.2}$	$\cdots$	$y_{.j}$	$\cdots$	$y_{.q}$	$y_{..}$

where: n_{ij} is the number of observations in the A_iB_j sub-class, $Y_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk}$, y_{ijk} is the kth observation in the ijth sub-class,

$$y_{ij} = Y_{ij}/n_{ij}, \quad Y_{i.} = \sum_{j=1}^{q} Y_{ij}, \quad Y_{.j} = \sum_{i=1}^{p} Y_{ij}, \quad n_{i.} = \sum_{j=1}^{q} n_{ij}$$

$$n_{.j} = \sum_{i=1}^{p} n_{ij}, \quad y_{i.} = Y_{i.}/n_{i.}, \quad y_{.j} = Y_{.j}/n_{.j},$$

$$Y_{..} = \sum_{i=1}^{p} Y_{i.} = \sum_{j=1}^{q} Y_{.j}, \quad n_{..} = \sum_{i=1}^{p} n_{i.} = \sum_{j=1}^{q} n_{.j}, \quad y_{..} = Y_{..}/n_{..}.$$

TABLE II

Preliminary Analysis of Variance

Source	d.f.	S.S.	M.S.	F
Between Sub-Classes	$pq - 1$	$\sum\limits_{i}^{p} \sum\limits_{j}^{q} \dfrac{Y_{ij}^2}{n_{ij}} - \dfrac{Y_{..}^2}{n_{..}}$	s_B^2	s_B^2/s_w^2
Within Sub-Classes	$n_{..} - pq$	Subtraction	s_w^2	
Total	$n_{..} - 1$	$\sum\limits_{i}^{p} \sum\limits_{j}^{q} \sum\limits_{k}^{n_{ij}} y_{ijk}^2 - \dfrac{Y_{..}^2}{n_{..}}$		

The optimum method of analysis of data of this type is the method of fitting constants from Yates [12]. Under the assumption of no interaction, a set of constants is fitted to the data so that the constants determine a set of sub-class means, with the property that the sum of weighted squares of the deviations of these means from the observed means is a minimum. If the classifications are large, this method becomes very tedious and laborious. In fact, one must write and solve at least $p + q$ normal equations, depending on which computational method is used.

Kendall [6] suggests using a much simpler and shorter, but less powerful test to compare the differences between the main effects. This method is known as the method of weighted squares of means. This incidentally, happens to be the optimum method if interaction is present. In this method unweighted marginal means are obtained.

$$(1) \qquad y'_{i.} = \frac{1}{q} \sum_{j=1}^{q} y_{ij}, \qquad y'_{.i} = \frac{1}{p} \sum_{i=1}^{p} y_{ij}$$

These are unbiased but inefficient estimates of the class means. By giving equal weight to all sub-class means, these marginal means for the A-classes are independent of the B-classes, and vice versa. The sums of squares due to the A-classes and B-classes are then calculated by (2).

$$(2) \qquad SS_A = \sum_{i=1}^{p} W'_{i.} y_{i.}'^2 - \left(\sum_{i=1}^{p} W'_{i.} y'_{i.} \right)^2 \Big/ \sum_{i=1}^{p} W'_{i.}$$

$$SS_B = \sum_{j=1}^{q} W'_{.j} y_{.j}'^2 - \left(\sum_{j=1}^{q} W'_{.j} y'_{.j} \right)^2 \Big/ \sum_{j=1}^{q} W'_{.j},$$

where $W'_{i.}$ and $W'_{.j}$ are given by (3).

(3) $$W'_{i.} = q^2 / \sum_{j=1}^{q} (1/n_{ij}), \qquad W'_{.j} = p^2 / \sum_{i=1}^{p} (1/n_{ij}).$$

$W'_{i.}$ and $W'_{.j}$ are the reciprocals of the variances (except for the factor σ^2) of $y'_{i.}$ and $y'_{.j}$. The mean squares of the A-classes and the B-classes are then tested with an F-test using the within sub-class mean square from the preliminary analysis. These are valid F tests but are low in power due to the inefficiency of the estimates $y'_{i.}$ and $y'_{.j}$ on which they are based.

A new method for this type of an analysis that is equally as simple to calculate as the method of weighted squares of means, but will be generally more powerful, will now be suggested.

For this new method, marginal means are obtained by (4).

(4) $$y''_{i.} = \frac{1}{n_{..}} \sum_{j=1}^{q} n_{.j} y_{ij}$$

$$y''_{.j} = \frac{1}{n_{..}} \sum_{i=1}^{p} n_{i.} y_{ij}$$

Since the weights $n_{.j}/n_{..}$ used in getting any mean, $y''_{i.}$, are independent of the row i concerned, the variation between the means $y''_{i.}$ is independent of the B-effects and vice versa as in the method of weighted squares of means.

This independence can be seen by considering the difference between any two A-class means, say $y''_{1.}$ and $y''_{2.}$. Since any $y_{ij} = m + a_i + b_j$, the marginal means are

(5) $$y''_{1.} = \frac{1}{n_{..}} \sum_{j=1}^{q} n_{.j}(m + a_1 + b_j)$$

$$= m + a_1 + \frac{1}{n_{..}} \sum_{j=1}^{q} b_j$$

$$y''_{2.} = \frac{1}{n_{..}} \sum_{j=1}^{q} n_{.j}(m + a_2 + b_j)$$

$$= m + a_2 + \frac{1}{n_{..}} \sum_{j=1}^{q} b_j .$$

Now $y''_{1.} - y''_{2.} = a_1 - a_2$, which shows that the variation between the means $y''_{i.}$ is independent of the B-effects.

It will be seen that the derivations of the marginal means $y_{i.}''$ and $y_{.j}''$ tend to give weight to sub-class means more in proportion to the numbers on which they are based than do the corresponding derivations of $y_{i.}'$ and $y_{.j}'$ in the method of weighted squares of means.

The sum of squares due to the A-classes and B-classes for the new proposed method are then obtained by (6).

$$(6) \qquad SS_A = \sum_{i=1}^{p} W_{i.}'' y_{i.}''^2 - \left(\sum_{i=1}^{p} W_{i.}'' y_{i.}'' \right)^2 / \sum_{i=1}^{p} W_{i.}''$$

$$SS_B = \sum_{j=1}^{q} W_{.j}'' y_{.j}''^2 - \left(\sum_{j=1}^{q} W_{.j}'' y_{.j}'' \right)^2 / \sum_{j=1}^{q} W_{.j}'' ,$$

where $W_{i.}''$ and $W_{.j}''$ are given by (7).

$$(7) \qquad W_{i.}'' = n_{..}^2 \Big/ \left(\sum_{j=1}^{q} n_{.j}^2 / n_{ij} \right), \qquad W_{.j}'' = n_{..}^2 \Big/ \left(\sum_{i=1}^{p} n_{i.}^2 / n_{ij} \right).$$

As in the method of the weighted squares of means, the weighting factors are the reciprocals of the variances (except for the factor σ^2) of the marginal means. The mean squares of the A-classes and the B-classes are then tested with an F-test using the within sub-class mean square from the preliminary analysis. These are also valid F-tests, but they are based on estimates of the class means, which will be generally more efficient and therefore more powerful than the F-tests obtained by the method of weighted squares of means.

This new method is proposed for cases in which p and q are large and when interaction can be assumed absent or negligible. If one is dubious about using the new method instead of the method of weighted squares of means, the following inequalities can be evaluated.

$$(8) \quad \sum_{i=1}^{p} W_{i.}'' - \sum_{i=1}^{p} W_{i.}''^2 / \sum_{i=1}^{p} W_{i.}'' > \sum_{i}^{p} W_{i.}' - \sum_{i=1}^{p} W_{i.}'^2 / \sum_{i=1}^{p} W_{i.}'$$

$$\sum_{j=1}^{q} W_{.j}'' - \sum_{j=1}^{q} W_{.j}''^2 / \sum_{j=1}^{q} W_{.j}'' > \sum_{j=1}^{q} W_{.j}' - \sum_{j=1}^{q} W_{.j}'^2 / \sum_{j=1}^{q} W_{.j}'$$

If these inequalities are satisfied the new method will be more powerful than the weighted squares of means.

The above inequalities which compare the method of weighted squares of means and the new proposed method have been obtained in the following way. Starting with the model $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$ we may write $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ where $\epsilon_{ij} = \sum_k \epsilon_{ijk} / n_{ij}$.

Dealing first with the method of weighted squares of means we have

(9) $$y'_{i.} = \sum_j y_{ij}/q$$

$$= \mu + \alpha_i + \sum_j \beta_j/q + \sum_j \epsilon_{ij}/q,$$

which reduces to

(10) $$y'_{i.} = \mu + \alpha_i + \epsilon'_{i.} \,,$$

where $\epsilon'_{i.} = \sum_j \epsilon_{ij}/q$ on applying the restriction $\sum_j \beta_j = 0$.
Now

(11) $$E(SS_A) = E \sum_i W'_{i.}(\alpha_i - \bar\alpha + \epsilon'_{i.} - \bar\epsilon)^2,$$

where

$$\bar\alpha = \sum_i W'_{i.}\alpha_i / \sum_i W'_{i.} \quad \text{and} \quad \bar\epsilon = \sum_i W'_{i.}\epsilon'_{i.} / \sum_i W'_{i.} \,.$$

Since the cross product terms are zero and since $W'_{i.}$ is the reciprocal of the variance of $\epsilon'_{i.}$ (except for the factor σ^2) this reduces to

(12) $$E(SS_A) = \sum_i W'_{i.}(\alpha_i - \bar\alpha)^2 + (p - 1)\sigma^2$$

Turning now to the proposed new method

(13) $$y''_{i.} = \sum_j n_{.j}y_{ij}/n_{..}$$

$$= \mu + \alpha_i + \sum_j n_{.j}\beta_j/n_{..} + \sum_j n_{.j}\epsilon_{ij}/n_{..}$$

which reduces to

(14) $$y''_{i.} = \mu + \alpha_i + \sum_j n_{.j}\beta_j/n_{..} + \epsilon''_{i.} \,,$$

where

$$\epsilon''_{i.} = \sum_j n_{.j}\epsilon_{ij}/n_{..} \,.$$

Now

(15) $$E(SS_A) = E \sum_i W''_{i.}(\alpha_i - \bar\alpha' + \epsilon''_{i.} - \bar\epsilon')^2,$$

where

$$\bar\alpha' = \sum_i W''_{i.}\alpha_i / \sum_i W''_{i.} \quad \text{and} \quad \bar\epsilon' = \sum_i W''_{i.}\epsilon''_{i.} / \sum_i W''_{i.} \,.$$

This reduces to

(16) $$E(SS_A) = \sum_i W''_{i.}(\alpha_i - \bar\alpha')^2 + (p - 1)\sigma^2.$$

From the expected values of the main effect sum of squares for the respective methods (12) and (16), it can be seen that the new method

will be more powerful for detecting real effects if

$$(17) \qquad \sum_i W''_{i.}(\alpha_i - \bar{\alpha}')^2 > \sum_i W'_{i.}(\alpha_i - \bar{\alpha})^2.$$

In the absence of knowledge of the α_i values it seems reasonable to replace each α_i^2 by a constant α^2 and each product term $a_i a_j$, $i \neq j$ by zero. On doing this the inequality (17) is readily seen to reduce to (8).

TABLE III

Data for Milk Yields, in Pounds, of Cows Freshening in Two Seasons

			Seasons	
			Fall and Winter	Spring and Summer
COWS	1	No.	2	2
		Total	969	632
		Mean	484.50	316.00
	2	No.	7	1
		Total	2477	262
		Mean	349.57	262.00
	3	No.	2	2
		Total	827	540
		Mean	413.50	270.00
	7	No.	2	4
		Total	572	890
		Mean	286.00	222.50
	9	No.	5	3
		Total	880	539
		Mean	176.00	179.67
	11	No.	7	4
		Total	2703	1280
		Mean	386.14	320.00
	21	No.	3	3
		Total	1184	1194
		Mean	394.67	398.00
	27	No.	4	5
		Total	1753	1701
		Mean	438.25	340.20

The quantities in (8) are also the coefficients of σ_A^2 and σ_B^2 in the expectations of the mean squares under Model II analysis of variance. Under the no interaction assumption and using the within sub-class mean square for testing, it is clear that the proposed method will be more sensitive if the coefficients of the "variance component" under test are larger, as was pointed out by a referee of this paper.

Numerical example: The following example was taken from pages 341–346, *Methods of Statistical Analysis* by C. H. Goulden [4].

Goulden gives the analysis of variance in Table IV for these data (Table III) by the method of fitting constants.

TABLE IV
Analysis of Variance (Fitting Constants)

Source	d.f.	S.S.	M.S.	F	5% Point
Cows	7	322,714	46,102	4.58	2.25
Seasons	1	58,177	58,177	5.78	4.08
Cows × Seasons	7	35,436	5,062		
Error	40	402,860	10,072		

It is seen that there is no evidence of an interaction, so the optimum method of completing the analysis would be by fitting constants which has been done.

Goulden also gives the analysis in Table V for these data by the method of weighted squares of means.

TABLE V
Analysis of Variance (Weighted Squares of Means)

Source	d.f.	S.S.	M.S.	F	5% Point
Cows	7	289,191	41,313	4.10	2.25
Seasons	1	64,810	64,810	6.43	4.08
Cows × Seasons	7	35,436	5,062		
Error	40	402,860	10,072		

It is seen that the method of weighted squares of means gives a sum of squares for cows 33,523 less than the method of fitting constants and a sum of squares for seasons 6,633 more than the method of fitting constants.

The weights used in the analysis are as follows:

$$W'_{1.} = 4.00000 \qquad W'_{6.} = 10.18184$$

$$W'_{2.} = 3.50000 \qquad W'_{7.} = 6.00000$$

$$W'_{3.} = 4.00000 \qquad W'_{8.} = 8.88888$$

$$W'_{4.} = 5.33332 \qquad W'_{.1} = 24.911936$$

$$W'_{5.} = 7.50000 \qquad W'_{.2} = 19.009920$$

Table VI gives the required calculations for obtaining the analysis of variance for the proposed new method.

Using the totals in Table VI we obtain the following sum of squares:

$$\text{Cows} = 5,776,003.795 - \frac{(16,533.135)^2}{49.99691} = 308,775$$

$$\text{Seasons} = 5,371,198.409 - \frac{(16,025.704)^2}{48.31119} = 55,180.$$

Comparing these sums of squares with the ones obtained by fitting constants, it is seen that for cows the new method is only 13,939 less and for seasons 2,997 less. Since there was no indication of an interaction, the proposed method gave results more near the results obtained by the method of fitting constants than did the method of weighted squares of means.

If we had used the inequalities to decide which method would be best before we did the analysis, we would have obtained the following values:

$$\text{Cows} \qquad 42.84540 > 42.36764$$

$$\text{Seasons} \quad 22.86865 > 21.564385.$$

The above inequalities tell us that the new proposed method will be more powerful than the method of weighted squares of means for both cow and season effects.

If we have a $p \times q$ table where p and q are both greater than 2, the only way we can obtain a test of interaction is by the method of fitting constants. If one does not wish a test of interaction or can assume that it is negligible, one can be sure he is obtaining the best approximation to the method of fitting constants by calculating the inequalities

TABLE VI
CALCULATIONS FOR PROPOSED NEW METHOD

		Fall and Winter	Spring and Summer	$n_{i.}$	$y''_{i.} = \dfrac{\sum_j n_{.j} y_{ij}}{n_{..}}$	$\sum_j n^2_{.j}/n_{ij}$	$W''_{i.} = \dfrac{n^2_{...}}{\sum_j n_{.j}/n_{ij}}$	$W''_{i.} y''_{i.}$	$W''_{i.} y''^2_{i.}$
C_1	No.	2	2	4					
	Mean	484.50	316.00		412.286		3.92000	1616.161	666,320.554
	$n^2_{.j}/n_{ij}$	512.000000	288.000000			800.000000			
	$n^2_{i.}/n_{ij}$	8.000000	8.000000						
C_2	No.	7	1	8					
	Mean	349.57	262.00		312.040		4.34177	1354.806	422,753.664
	$n^2_{.j}/n_{ij}$	146.285714	576.000000			722.285714			
	$n^2_{i.}/n_{ij}$	9.142857	64.000000						
C_3	No.	2	2	4					
	Mean	413.50	270.00		352.000		3.92000	1379.840	485,703.680
	$n^2_{.j}/n_{ij}$	512.000000	288.000000			800.000000			
	$n^2_{i.}/n_{ij}$	8.000000	8.000000						
C_7	No.	2	4	6					
	Mean	286.00	222.50		258.786		4.78049	1237.124	320,150.371
	$n^2_{.j}/n_{ij}$	512.000000	144.000000			656.000000			
	$n^2_{i.}/n_{ij}$	18.000000	9.000000						
C_9	No.	5	3	8					
	Mean	176.00	179.67		177.573		7.90322	1403.398	249,205.593
	$n^2_{.j}/n_i$	204.800000	192.000000			396.800000			

				Totals			
C_{11}	No.	7	4	11			
	Mean	386.14	320.00	357.794	10.80315	3865.302	1,382,981.864
	$n^2_{.j}/n_{ij}$	146.285714	144.000000	290.285714			
	$n^2_{i.}/n_{ij}$	17.285714	30.250000				
C_{21}	No.	3	3	6			
	Mean	394.67	398.00	396.097	5.88000	2329.050	922,529.718
	$n^2_{.j}/n_{ij}$	341.333333	192.000000	533.333333			
	$n^2_{i.}/n_{ij}$	12.000000	12.000000				
C_{27}	No.	4	5	9			
	Mean	438.25	340.20	396.229	8.44828	3347.454	1,326,358.351
	$n^2_{.j}/n_{ij}$	256.000000	115.200000	371.200000			
	$n^2_{i.}/n_{ij}$	20.250000	16.200000				
	Totals				49.99691	16,533.135	5,776,003.795
	$n_{.j}$	32	24	56			
$y''_{.j} = (\sum_i n_{i.} y_{ij})/n_{..}$		358.435	288.967				
$\sum_i n^2_{i.}/n_{ij}$		105.478571	168.783333				
$W''_{.j} = n^2_{..}/(\sum_i n^2_{i.}/n_{ij})$		29.73116	18.58003	48.31119			
$W''_{.j}\, y''_{.j}$		10,656.688	5,369.016	16,025.704			
$W''_{.j}\, y''^2_{.j}$		3,819,729.963	1,551,468.446	5,371,198.409			

given by (8) as was done above and then decide on the method of analysis.

REFERENCES

1. Anderson, R. L. and Bancroft, T. A. *Statistical Theory in Research.* McGraw-Hill Book Co., Inc., New York, 1952.
2. Brandt, A. E. The Analysis of Variance in a $2 \times s$ Table With Disproportionate Frequencies. *J. A. S. A., 28*, 1933, pp. 164–173.
3. Crump, S. L. The Estimation of Variance Components in Analysis of Variance. *Biometrics Bulletin, 2*, 1946, pp. 7–11.
4. Goulden, C. H. *Methods of Statistical Analysis.* John Wiley and Sons, Inc., New York, 1952.
5. Hazel, L. N. The Covariance Analysis of Multiple Classification Tables with Unequal Sub-Class Numbers. *Biometrics Bulletin, 2*, 1946, pp. 21–25.
6. Kendall, M. G. *The Advanced Theory of Statistics.* Vol. II, Third Edition, Hafner Publishing Co., New York, 1951.
7. Rao, C. R. *Advanced Statistical Methods in Biometric Research.* John Wiley and Sons, Inc., New York, 1952.
8. Snedecor, G. W. *Statistical Methods.* The Iowa State College Press, Ames, Iowa, 1950.
9. Snedecor, G. W. and Cox, G. M. Disproportionate Sub-class Numbers in Tables of Multiple Classification. *Research Bulletin No. 180*, Agricultural Experiment Station, Iowa State College of Agriculture and Mechanic Arts, Ames, Iowa, 1935.
10. Stevens, W. L. Statistical Analysis of a Non-Orthogonal Tri-Factorial Experiment. *Biometrika, 35*, 1948, pp. 346–367.
11. Wilks, S. S. *Mathematical Statistics.* Princeton University Press, Princeton, N. J., 1950.
12. Yates, F. The Principles of Orthogonality and Confounding in Replicated Experiments. *Journal of Agricultural Science, 23*, 1933, pp. 108–145.
13. Yates, F. The Analysis of Multiple Classifications With Unequal Numbers in the Different Classes. *J. A. S. A., 29*, 1934, pp. 51–66.

A METHOD OF ANALYSIS FOR A DOUBLE CLASSIFICATION ARRANGED IN A TRIANGULAR TABLE

Nell Ditchburne

Division of Mathematical Statistics, C.S.I.R.O., Melbourne

I. INTRODUCTION

When an experiment is designed to test the effect of two factors each at several levels, on some measurable quantity, the data may be arranged in a two-way table. When each factor is tested at all levels of the other factor, and the number of observations in the subclasses are equal or proportional, estimates of the effects of each factor are easily obtained, the analysis of variance of the data is simple, and interpretation of results straightforward, as the data are orthogonal. However, when the numbers in the subclasses are unequal, or when all levels of one factor are not tested at all levels of the other, so that some subclasses are completely missing, the data are non-orthogonal, and a method of analysis must be found to suit the design of the particular experiment.

The method of analysis described in this paper is suitable for use with data which may be arranged in a triangular table such as that set out diagrammatically below. In the diagram, data are available for the subclasses marked x.

Rows	Columns			
	1	2	3	4
1	x	x	x	x
2	x	x	x	
3	x	x		
4	x			

The triangular type of design can often occur in experiments in which the levels of one factor may be found from the sum or difference of two other factors. For instance, in certain experiments in animal production, the effects of age at mating, year dropped, and year mated on percentage reproduction are investigated. In this case, age at mating is equal to the difference between year mated and year dropped. The data could be arranged in triangular tables, taking two factors at a time, that is, year mated and year dropped, age at mating and year dropped, or age at mating and year mated.

In chemical work, the effect of varying temperature range, with differing initial and final temperatures, may be analysed by this method; thus the effects of any two of range, initial temperature and final temperature may be determined. This triangular type of arrangement was employed in an experiment on the chemical retting of flax; in this case the effects of range and final temperature were considered, because rets were held longer at the final temperatures than at initial temperatures. A numerical example from this experiment is given in Section IV.

An analysis of a table of this type is possible only on the assumption that interactions do not exist, that is, that the effects of the two factors are additive. On this assumption, the method of fitting constants by least squares, as described by Yates (1933), is appropriate for all experiments with multiple classifications, whether or not there are empty subclasses. By this method of analysis, the correctness of the assumption of non-existent interactions may be tested; if there is evidence that interactions do exist, the data must be reanalysed, to examine separately the effects of each factor. For the two examples quoted above, the method of analysis to be described would be appropriate for those modes of classification of the data which gave effects which were additive.

Data arranged in a triangular table may be analysed without solving least squares equations, the sum of squares for each factor, freed from the effects of the other factor, being obtained from a set of orthogonal comparisons. This method, which is the one described in this paper, is more rapid than that involving the solution of least squares equations. However, it is applicable only under certain conditions: either the number of replicates in each subclass of the table should be equal, which is the general case discussed in Section III, or the number of replicates in each subclass within a single level of one of the factors should be equal. In the latter case, the table of subclass numbers would be of the form

n_1	n_2	n_3	n_4
n_1	n_2	n_3	
n_1	n_2		
n_1			

or

n_1	n_1	n_1	n_1
n_2	n_2	n_2	
n_3	n_3		
n_4			

II. ORTHOGONAL COMPONENTS OF A SUM OF SQUARES

Cochran and Cox (1950) give a summary of the conditions under which comparisons among k treatment totals T_i are orthogonal. If the number of replicates of the ith treatment is n_i, then the function

$$Z_w = l_{w1}T_1 + l_{w2}T_2 + \cdots l_{wk}T_k$$

is a comparison among the T_i provided

$$n_1 l_{w1} + n_2 l_{w2} + \cdots n_k l_{wk} = 0.$$

Two comparisons are orthogonal if

$$n_1 l_{11} l_{21} + n_2 l_{12} l_{22} + \cdots n_k l_{1k} l_{2k} = 0.$$

The quantity Z_w^2/D_w, where

$$D_w = n_1 l_{w1}^2 + n_2 l_{w2}^2 + \cdots + n_k l_{wk}^2$$

is a component of the sum of squares for treatments, and has one degree of freedom.

Among the k treatments, there are $k - 1$ comparisons Z_w which are orthogonal, and therefore it follows that

$$\sum (Z_w^2/D_w) = \sum (T_i^2/n_i) - (\sum T_i)^2 / \sum n_i .$$

III. METHOD OF ANALYSIS FOR TRIANGULAR CLASSIFICATIONS

A description of the analysis for a triangular design when all the numbers in all subclasses are equal to n will now be given.

If k is the number of rows and columns of the table, there are $k - 1$ degrees of freedom in the sum of squares for the main effect of either rows or columns; each of these sums of squares may be subdivided into $k - 1$ orthogonal components.

If the totals for the ith row and column be R_i and C_i respectively, then it may be shown that the quantities

$$Y_i = (i - 1)R_i - \sum_{j=1}^{i-1} R_j + \sum_{j=k-i+2}^{k} C_j \qquad (i = 2, 3 \cdots k)$$

are mutually orthogonal comparisons among the rows.

The quantity Y_i is made up of $n(k - i + 1)$ values each multiplied by $i - 1$, less $n(i - 1)(k - i + 1)$ other values. Hence the variance of Y_i is proportional to

$$n[(i - 1)^2(k - i + 1) + (i - 1)(k - i + 1)] = ni(i - 1)(k - i + 1)$$

The sum of squares for rows freed from column effects is therefore given by

$$\frac{1}{n} \sum_{i=2}^{k} Y_i^2 / i(i - 1)(k - i + 1).$$

The sum of squares for columns, freed from row effects, is obtained similarly.

The sum of squares for rows, column effects being ignored, is

$$\frac{1}{n} \left[\sum_{i=1}^{k} \left(R_i^2/(k - i + 1) \right) - \left(\sum_{i=1}^{k} R_i \right)^2 / \tfrac{1}{2}(k^2 + k) \right]$$

and that for columns, row effects being ignored, is obtained in a similar manner.

The interaction sum of squares has $\frac{1}{2}(k - 1)(k - 2)$ degrees of freedom, and is obtained from the total sum of squares between sub-classes minus the sums of squares for rows with columns eliminated and columns ignoring rows.

Estimates of the means for each factor, freed from effects of the other, may be obtained in the following way.

If the estimates for the ith row are r_i, then

$$r_2 - r_1 = Y_2/n(k - 1),$$

$$2r_3 - r_2 - r_1 = Y_3/n(k - 2),$$

and generally

$$(i - 1)r_i - \sum_{j=1}^{i-1} r_j = Y_i/n(k - i + 1);$$

also

$$kr_1 + (k - 1)r_2 + (k - 2)r_3 + \cdots r_k = 0.$$

The solutions of these equations are

$$r_1 = -\frac{1}{nk(k + 1)} \sum_{j=2}^{k} \frac{Y_j(k + j)}{j(j - 1)}$$

$$r_2 = \frac{1}{n} \left[\frac{Y_2}{(k - 1)} - \frac{1}{k(k + 1)} \sum_{j=2}^{k} \frac{Y_j(k + j)}{j(j - 1)} \right]$$

etc.

Generally,

$$r_i = \frac{1}{n} \left[\frac{Y_i}{(k - i + 1)(i - 1)} + \sum_{j=2}^{i-1} \frac{Y_i}{j(k - j + 1)(j - 1)} \right.$$

$$\left. - \frac{1}{k(k + 1)} \sum_{j=2}^{k} \frac{Y_i(k + j)}{j(j - 1)} \right]$$

$$= \frac{1}{n} \left[\frac{Y_i}{(k - i + 1)(i - 1)} + \frac{1}{k(k + 1)} \left(\sum_{j=2}^{i-1} \frac{Y_i}{k - j + 1} \right. \right.$$

$$\left. \left. - \sum_{j=1}^{k} \frac{Y_i(k + j)}{j(j - 1)} \right) \right]$$

The difference between two estimates r_i and r_h, $(i > h)$, is

$$\frac{1}{n} \left[\frac{Y_i}{(i - 1)(k - i + 1)} - \frac{Y_h}{h(k - h + 1)} + \sum_{j=h+1}^{i-1} \frac{Y_i}{j(j - 1)(k - j + 1)} \right]$$

As the Y_i are orthogonal, the variance of $r_i - r_h$ is simply

$$\frac{1}{n^2} \left[\frac{V(Y_i)}{[(i - 1)(k - i + 1)]^2} + \frac{V(Y_h)}{[h(k - h + 1)]^2} \right.$$

$$\left. + \sum_{j=h+1}^{i-1} \frac{V(Y_i)}{[j(j - 1)(k - j + 1)]^2} \right]$$

Since the variance of Y_i is proportional to

$$ni(i - 1)(k - i + 1),$$

it follows that $V(r_i - r_h)$ is proportional to

$$\frac{1}{n} \left[\frac{i(i - 1)(k - i + 1)}{[(i - 1)(k - i + 1)]^2} + \frac{h(h - 1)(k - h + 1)}{[h(k - h + 1)]^2} \right.$$

$$\left. + \sum_{j=h+1}^{i-1} \frac{j(j - 1)(k - j + 1)}{j(k - j + 1)(j - 1)^2} \right]$$

$$= \frac{1}{n} \left[\frac{i}{(i - 1)(k - i + 1)} + \frac{h - 1}{h(k - h + 1)} \right.$$

$$\left. + \sum_{j=h+1}^{i-1} \frac{1}{j(k - j + 1)(j - 1)} \right].$$

Estimates of column means c_i, and variances of differences between two estimates c_i and c_h may be obtained in a similar manner.

<center>IV. NUMERICAL EXAMPLE</center>

The data presented in Table 1 are values of buffer capacity after retting for 99 hours, for four varieties of flax, A, B, C and D,

TABLE 1
Individual Values of Buffer Capacity.

Temp.	Variety	Range				Totals for Ranges			
		0	4	8	12	0 to 12	0 to 8	0 to 4	0
40	A	7.42	6.69	6.40	5.84				
	B	7.92	9.90	6.43	9.33				
	C	7.61	6.18	9.53	6.81				
	D	7.54	6.84	7.14	6.37				
	Total	30.49	29.61	29.50	28.35	117.95	89.60	60.10	30.49
36	A	6.60	5.68	5.85	4.96				
	B	7.58	6.82	7.32	5.72				
	C	6.45	6.40	5.97	5.71				
	D	6.55	7.26	6.42	6.00				
	Total	27.18	26.16	25.56	22.39	101.29	78.90	53.34	27.18
32	A	5.21	4.76	4.89					
	B	6.74	6.31	5.47					
	C	5.98	5.98	4.54					
	D	6.52	6.01	5.41					
	Total	24.45	23.06	20.31			67.82	47.51	24.45
28	A	3.80	3.88						
	B	5.22	5.20						
	C	4.50	4.18						
	D	4.64	4.38						
	Total	18.16	17.64					35.80	18.16

24	A	2.89						74.87
	B	3.25						93.21
	C	2.90						82.74
	D	3.63						84.71
	Total	12.67					12.67	335.53
Totals (40 to 24)	A							
	B							
	C							
	D							
	Total							
40 to 24	all	112.95	96.47	75.37	50.74			
40 to 28		100.28	78.83	55.06				
40 to 32		82.12	55.77					
40 to 36		57.67						

four temperature ranges, 0, 4, 8 and 12 degrees Centigrade, and five final temperatures, 24, 28, 32, 36 and 40 degrees Centigrade. It is seen that this example differs slightly from the general case described in Section III, but a similar method of analysis is used. Before analysis, data should be arranged so that the table is of the form shown in section I; otherwise the formulae of section III are meaningless.

<div align="center">

TABLE 2

Buffer Capacity at 99 hours Retting.

Analysis of Variance

</div>

	D.F.	Sum of Squares		Mean Square
		Range (T elim.)	Temp. (R elim.)	
Varieties	3	12.1584	12.1584	4.0528**
Range	3	*4.6622*	4.3798	1.5541*
Final Temperature	4	84.0110	*84.2934*	21.0734**
$R \times T$	6	1.3715	1.3715	0.2286(n)
Error	39	20.1167	20.1167	0.5158
Total	55	122.3198	122.3198	

The analysis of variance is shown in Table 2. A detailed description of the methods used in computing the required sums of squares and estimates is given below, together with various short methods of computation, which are preferable to direct use of the formulae of section III.

Sums of Squares

Varieties. As all temperatures and ranges are equally represented in all varieties, the sum of squares is obtained direct from the variety totals.

$$(74.87^2 + 93.21^2 + 82.74^2 + 84.71^2)/14 - (335.53^2)/56 = 12.1584$$

Range (a) unadjusted for temperature effect.

This sum of squares is obtained by the method generally employed when subclass numbers are unequal:

$$(112.95^2/20 + 96.47^2/16 + 75.37^2/12 + 50.74^2/8 - 335.53^2/56)$$

$$= 4.3798$$

(b) with temperature effect eliminated.

To simplify computation of the values $Y_,$, certain extra totals, shown in Table 1, are required. For instance, for comparison of ranges 0 and

4, a sum of squares with one degree of freedom is obtained as

$$(96.47 - 112.95 + 12.67)^2/(2 \times 16).$$

This is a direct comparison of ranges 0 and 4, summed for temperatures 40 to 28:

$$(96.47 - 100.28)^2/(2 \times 16) = (-3.81)^2/32$$

$$= 0.4536.$$

From the totals for ranges 0, 4 and 8, summed for temperatures 40 to 32, a sum of squares is obtained for the comparison of the mean for ranges 0 and 4 with the mean for range 8:

$$(2 \times 75.37 - 82.12 - 78.83)^2/(6 \times 12) = (-10.21)^2/72$$

$$= 1.4478.$$

From the totals for ranges 0 to 12, summed for temperatures 40 and 36, the sum of squares for the comparison of the mean for range 12 with the mean for ranges 0, 4 and 8 is

$$(3 \times 50.74 - 57.67 - 55.77 - 55.06)^2/(12 \times 8) = (-16.28)^2/96$$

$$= 2.7608.$$

By the addition of these separate squares, we obtain the value of 4.6622, with 3 degrees of freedom, which is shown in Table 2.

Temperature (a) unadjusted for range effect.

The sum of squares, with 4 degrees of freedom, is

$$(117.95^2/16 + 101.29^2/16 + 67.82^2/12 + 35.80^2/8 + 12.67^2/4$$

$$- 335.53^2/56) = 84.0110.$$

(b) with effect of range eliminated.

The difference between means for temperatures 40 and 36 is free of any range effect; the sum of squares, with 1 degree of freedom, is

$$(101.29 - 117.95)^2/(2 \times 16) = (-16.66)^2/32$$

$$= 8.6736.$$

For ranges 0 to 8, the sum of squares for the comparison of the mean for temperatures 40 and 36 with the mean for temperature 32 is

$$(2 \times 67.82 - 101.29 - 117.95 + 50.74)^2/(6 \times 12)$$

$$= (2 \times 67.82 - 89.60 - 78.90)^2/(6 \times 12)$$

$$= (-32.86)^2/72$$

$$= 14.9969.$$

For ranges 0 to 4, the sum of squares for the comparison of the mean for temperatures 40, 36 and 32 with the mean for temperature 28 is

$$(3 \times 35.80 - 60.10 - 53.34 - 47.51)^2/(12 \times 8) = (-53.55)^2/96$$
$$= 29.8709.$$

For range 0, the sum of squares for comparison of the mean of temperatures 40 to 28 with the mean for temperature 24 is

$$(4 \times 12.67 - 30.49 - 27.18 - 24.45 - 18.16)^2/(20 \times 4)$$
$$= (-49.60)^2/80$$
$$= 30.7520.$$

The total sum of squares for temperatures, with range effect eliminated, is therefore

$$8.6736 + 14.9969 + 29.8709 + 30.7520$$
$$= 84.2934, \text{ with 4 degrees of freedom.}$$

A check for these sums of squares is that

Range unadjusted + temperature adjusted
= Range adjusted + temperature unadjusted,
i.e., 4.3798 + 84.2934 = 4.6622 + 84.0110.

The multipliers used in obtaining the orthogonal comparisons are summarized in Table 3; from the Table it is easily verified that the comparisons satisfy the conditions for orthogonality given in Section II.

Interaction of Range and Temperature

This sum of squares is obtained from the sum of squares for subclasses minus the sums of squares for temperature adjusted and range unadjusted, or temperature unadjusted and range adjusted.

$$(30.49^2 + 29.61^2 + \cdots 12.67^2)/4 - 335.53^2/56 - 84.2934 - 4.3798$$
$$\text{or} \quad - 84.0110 - 4.6622.$$

The interaction of varieties with other effects is used as an estimate of error for testing the significance of the effects of range and temperature and their interaction, as it is considered that the four varieties are replications of the experiment.

The interaction effect of range and temperature is not significant; this confirms the original hypothesis, and indicates that the analysis by this method gives a valid test of the significance of the effects of range and temperature.

TABLE 3

Multipliers used in obtaining orthogonal comparisons.

Range Effect

	n_i	l_i Range					Divisor for Sum of Squares
		0	4	8	12	24	
Totals over temps.							
40 to 28	16	−1	+1				32
40 to 32	12	−1	−1	+2			72
40 to 36	8	−1	−1	−1	+3		96

Temperature effect

	n_i	Temperature					Divisor for Sum of Squares
		40	36	32	28	24	
Totals over ranges							
0	4	−1	−1	−1	−1	+4	80
0 to 4	8	−1	−1	−1	+3		96
0 to 8	12	−1	−1	+2			72
0 to 12	16	−1	+1				32

Estimates

As this example differs from the general case, estimates are more easily obtained by solution of equations using the Y_i which were obtained in calculation of sums of squares than by adjustment of the general solution for the r_i given in Section III.

The equations to be solved are

$$-t_{40} + t_{36} \qquad\qquad\qquad\qquad = -16.66/16$$

$$-t_{40} - t_{36} + 2t_{32} \qquad\qquad\qquad = -32.86/12$$

$$-t_{40} - t_{36} - t_{32} + 3t_{28} \qquad\qquad = -53.55/8$$

$$-t_{40} - t_{36} - t_{32} - t_{28} + 4t_{24} = -49.60/4$$

$$4t_{40} + 4t_{36} + 3t_{32} + 2t_{28} + t_{24} = 0$$

and

$$r_0 + r_4 \qquad\qquad\qquad = -3.81/16$$

$$-r_0 - r_4 + 2r_8 \qquad\qquad = -10.21/12$$

$$-r_0 - r_4 - r_8 + 3r_{12} = -16.28/8$$

$$5r_0 + 4r_4 + 3r_8 + 2r_{12} = 0.$$

The solutions to these equations are identical with those obtained from the normal least squares equations,

$$4(4t_{40} + r_0 + r_4 + r_8 + r_{12}) = T_{40}$$

etc.

The grand mean is added to the solutions to give the following values:

$$t_{40} = 7.48 \qquad\qquad r_0 = 6.31$$

$$t_{36} = 6.44 \qquad\qquad r_4 = 6.07$$

$$t_{32} = 5.59 \qquad\qquad r_8 = 5.77$$

$$t_{28} = 4.28 \qquad\qquad r_{12} = 5.37$$

$$t_{24} = 6.31$$

REFERENCES

Cochran, W. G., and Cox, G. M. *Experimental Designs.* John Wiley & Sons, 1950.

Yates, F. (1933). The principles of orthogonality and confounding in replicated experiments. *J. Agric. Sci.*, *23*: 108.

APPROPRIATE SCORES IN BIO-ASSAYS USING DEATH-TIMES AND SURVIVOR SYMPTOMS*

JOHANNES IPSEN

Associate Professor of Public Health, Harvard School of Public Health

Introduction

Death and survival are the most commonly used markers in biological standardization and in evaluation of medical therapy. Numerous methods are available for estimating the parameters of the dosage-mortality curve, and the biologist is often in a state of embarrassment of riches in the choice between a dozen well recommended methods to determine his LD50, ED50, TCiD50, or other "D50's" with which he is concerned.

The experimental records of good biologists and clinicians usually contain more information than death or survival of the subjects on given treatments, but either the "LD50-fixation" of the investigator prevents these data from entering in the evaluating process, or they are left out because no simple method is available to utilize these observations. In many assays, individual deaths have a quantitative connotation in terms of the time period from exposure to the lethal agent until death occurs. Survival has also quantitative aspects such as time of recovery, severity of symptoms at the time when death is no longer expected, etc.

Some data may consist mostly of deaths occurring at different times, in other experiments survivors are in excess. In these cases, group mortality percentages are too large or too small, respectively, to be of biometric use. Attempts are then made to find a transformation of death times or survivor symptoms that can be used as response meta-meter with approximately linear relationship to dose or with approximately normal distribution, or both. However, in such attempts the problem of truncated or censured distribution will sooner or later present itself, leaving the investigator either in uncomfortable indecision or involved in excessive computation.

The most complex situation arises when some treatments of a series of graded treatments present groups with total mortality, and some present total survivorship, and intermediate treatments have partial mortality. The biologist, feeling that restriction to mortality percentages will throw some expensive observations out, may attempt to combine his graded observations on deaths and survivors in one continuous score system to include all data in the estimate of treatment effect. This writer has tried some graphical methods with the purpose of obtaining a normal distribution of such continuous score system in clinical (1) and laboratory data (2). In some immunological bio-assays a solution was attempted by assigning a score system (3), which—by graphical trial and error—gave a linear dose response curve. Both methods were unsatisfactory because there was no available method to show that the score systems utilized the biological data efficiently, and because the graphical methods were too subjective for reproduction by other workers.

Appropriate Scores for Linear Bio-Assays

The method to be presented provides a set of scores for multiple, mutually exclusive observations that satisfies *one* criterion for an efficient bio-assay:

The variance of the linear regression of the mean scores on log dose is the highest possible fraction of the total variance of a set of graded dose experiments.

Let us first consider an experiment where k different logarithmic doses x_1 , x_2 $\cdots$ x_k of the same preparation are given to k groups containing $a_{1.}, a_{2.} \cdots a_{k.}$ individuals. The observations are classes in m categories, e.g., ranging in death at 1, 2, 3 days and in severity of symptoms at survival. Subdivision in categories will be directed by biological experience and by the frequency of observation in each category.

If the number of individuals treated with log dose x_i that are observed in category j, is designated a_{ij} , the data may be arranged as follows (Table 1).

We shall use the sum of squares of deviations of the doses

$$S_{xx} = \sum_i x_i^2 . a_{i.} - (\sum_i x_i a_{i.})^2 / a_{..} \tag{1}$$

and the normalized individual sums of products for each category

$$d_j = \sum_i x_i a_{ij} - \frac{a_{.j} \sum_i x_i a_{i.}}{a_{..}} \tag{2}$$

TABLE 1

log dose	Reaction category						Number Individuals per dose
	1	2	$\cdots$	j	$\cdots$	m	
x_1	a_{11}	a_{12}	$\cdots$	a_{1j}	$\cdots$	a_{1m}	$a_{1.}$
x_2	a_{21}	a_{22}	$\cdots$	a_{2j}	$\cdots$	a_{2m}	$a_{2.}$
.	.	.	$\cdots$	.	$\cdots$	.	.
.	.	.	$\cdots$	.	$\cdots$	.	.
x_i	a_{i1}	a_{i2}	$\cdots$	a_{ij}	$\cdots$	a_{im}	$a_{i.}$
.	.	.	$\cdots$	.	$\cdots$	.	.
.	.	.	$\cdots$	.	$\cdots$	.	.
x_k	a_{k1}	a_{k2}	$\cdots$	a_{ki}	$\cdots$	a_{km}	$a_{k.}$
Number per Category	$a_{.1}$	$a_{.2}$	$\cdots$	$a_{.j}$	$\cdots$	$a_{.m}$	Total $= a_{..}$

It is evident that

$$\sum_j d_j = 0 \tag{3}$$

If each category is given a score or measurement z_1, z_2, $\cdots z_j$, $\cdots$, z_m belonging to a system (Z) we obtain the total sum of squares

$$S_{zz} = \sum_j a_{.j} z_j^2 - \left(\sum_j a_{.j} z_j \right)^2 / a_{..} \tag{4}$$

and the sum of products

$$S_{zz} = \sum_j (d_j z_j) \tag{5}$$

The problem is to find a system (C), c_1, $\cdots$, c_i, $\cdots$, c_m which maximizes the function

$$\text{Max } \{r_{zz}^2\} = \text{Max } \left\{ \frac{(S_{zz})^2}{S_{xx} S_{zz}} \right\} = \frac{(S_{cx})^2}{S_{xx} S_{cc}} \tag{6}$$

It will be shown that such a system is obtained if

$$c_i = \frac{d_i}{a_{.i}} \tag{7}$$

or any linear transformation of c_i;

Proof:

Two $(m - 1)$ $(m - 1)$ matrices are arranged as follows:

$$
A = \begin{bmatrix}
a_{.2} - \dfrac{a_{.2}^2}{a_{..}} & - \dfrac{a_{.3}a_{.2}}{a_{..}} & \cdots & - \dfrac{a_{.m}a_{.2}}{a_{..}} \\[2ex]
- \dfrac{a_{.2}a_{.3}}{a_{..}} & a_{.3} - \dfrac{a_{.3}^2}{a_{..}} & \cdots & - \dfrac{a_{.m}a_{.3}}{a_{..}} \\[2ex]
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\[1ex]
- \dfrac{a_{.2}a_{.m}}{a_{..}} & - \dfrac{a_{.3}a_{.m}}{a_{..}} & \cdots & a_{.m} - \dfrac{a_{.m}^2}{a_{..}}
\end{bmatrix}
$$

$$
D = \frac{1}{S_{xx}}
\begin{bmatrix}
d_2^2 & d_2 d_3 & \cdots & d_m d_2 \\
d_2 d_3 & d_3^2 & \cdots & d_m d_3 \\
\cdot & \cdot & \cdot\,\cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
d_2 d_m & d_3 d_m & \cdots & d_m^2
\end{bmatrix}
$$

A third matrix is constructed where each element is θ times the elements of the A-matrix minus the corresponding elements of the D-matrix. Equating the determinant of this matrix with zero, the following explicit solution is obtained:

$$
\theta = \frac{1}{S_{xx}} \sum_i \frac{d_i^2}{a_{.i}}
\tag{8}
$$

We can now compute a score system (c') where $c_1' = 0$ and $c_m' = 1$ by solving $(m - 2)$ sets of equations of the form

$$
-c_2'\left(\frac{a_{.2}a_{.i}}{a_{..}}\,\theta + \frac{d_2 d_i}{S_{xx}}\right) - \cdots + c_i'\left[\left(a_{.i} - \frac{a_{.i}^2}{a_{..}}\right)\theta - \frac{d_i^2}{S_{xx}}\right] - \cdots
$$

$$
- c_{m-1}'\left(\frac{a_{.i}a_{m-1}}{a_{..}}\,\theta + \frac{d_i d_{m-1}}{S_{xx}}\right) = \frac{a_{.i}a_{.m}}{a_{..}}\,\theta + \frac{d_i d_m}{S_{xx}}
$$

The solutions are:

$$
c_i' = \frac{1}{\dfrac{d_m}{a_{.m}} - \dfrac{d_1}{a_{.1}}}\left(\frac{d_i}{a_{.i}} - \frac{d_1}{a_{.1}}\right)
\tag{9}
$$

It is clear that the (C') system is linear with the (C) system as defined in (7) and hence *we can use $c_i = d_i/a._i$ as the appropriate score system.*

Inserting (7) in (4) and (5) we obtain

$$S_{cx} = S_{cc} = \sum_i d_i^2/a._i = \theta \cdot S_{xx} \qquad (10)$$

and from (6)

$$\text{Max } \{r_{xx}^2\} = \frac{(S_{cx})^2}{S_{xx}S_{cc}} = \theta$$

Significance of θ

The square root of θ is the correlation coefficient between dose and scores and hence a significance test of the information extracted from a given experiment in respect to dose can be had from a table of r, entered with $(a.. - m)$ degrees of freedom.

Another test which is somewhat simpler, because it does not involve extracting a square root consists in assuming $a.. \theta$ to follow a χ^2-distribution with $m - 1$ degrees of freedom. This test is justified since $a.. \theta$ is the linear component in x of the χ^2 for an $(m) \times (k)$ contingency table.

We shall in the following use the term

$$\chi_c^2 = \theta a.. \qquad (11)$$

Example 1. Clinical scoring of typhus fever.

Ecke et al. (4) found that persons immunized to varying degrees with typhus vaccine responded to subsequent typhus fever in various categories $A, B, \cdots, F$ classified in order of severity of clinical symptoms with F corresponding to death. The authors scored the six categories 0, 20, 40, 60, 80 and 100 with $A = 0$ and $F = 100$ and computed mean scores for each degree of immunization. These mean scores showed a good correlation with degree of immunization. (Table 2)

One may ask the questions: 1. What is the maximum information that can be obtained considering degree of immunization a linearly progressive scale? 2. What is the appropriate score system for the six categories under these conditions?

The cases were observed among the personnel in the typhus wards in Cairo during World War II. Immunization was attempted but some contracted the disease before completion of a three course dose schedule. The clinical categories were defined before the records were scrutinized.

Table 2 presents the data with "dose" ranked in equidistant classes $1 - 5$ (x). The computed appropriate scores that will form the best

TABLE 2

Example 1

Clinical severity of 71 cases of typhus fever related to degree of previous immunization
(Ecke, et al. (4))

Degree of previous typhus immunization	x_i	Number of persons with typhus in clinical category (a_{ij})						Total $a_i.$	Mean Score	
		A	B	C	D	E	F		Optimal c'	Assumed z
None	1			5	2		3	10	75.1	62.0
1 Dose $<$ 12 days	2		4	8	4		1	17	57.0	43.5
1 Dose $>$ 12 days	3		2	8	1			11	52.9	38.2
2 Doses	4		4	3				7	35.0	28.6
3 Doses	5	1	20	5				26	24.8	23.1
$a_{.j}$		1	30	29	7	0	4	71 $(a_{..})$		
$\sum x_i a_{ij}$		1	130	82	13		5	235	$S_{xx} = \dfrac{11444}{71}$	
d_j (eq. 2)		$\dfrac{120}{71}$	$\dfrac{2180}{71}$	$\dfrac{-993}{71}$	$\dfrac{-722}{71}$		$\dfrac{-585}{71}$			
c_j (eq. 7)		1.69	1.02	-0.48	-1.45		-2.06			
$c'_j = \dfrac{c_j - c_A}{c_F - c_A} 100$		0	17.8	57.9	83.8		100			
z_j (assumed)		0	20	40	60		100			

(eq. 8) $\theta = 0.45148$ $r_{xc} = 0.672$ (D.F. $= 66$) $P < .001$
$a_{..}\theta = \chi^2_c = 32.055$ (D.F. $= 4$) $P < .001$

linear regression on x are shown in the row marked c_j. These scores
are transformed linearly to c'_j so that direct comparison with the assumed
scores z_j is facilitated.

The maximum information is highly significant whether the r-test
or the χ^2-test is applied.

The appropriate scores are somewhat different from the assumed
clinical score. As so often happens in medical science the death category
was assigned a higher relative weight than the adjacent survivor cate-
gories.

A third and more interesting question is whether the assumed score

system was significantly less informative than the appropriate score system. This question needs a few theoretical considerations.

Test for "lost information".

Since the term $a_{..}\theta$ represents maximum information under the specified conditions, and

$$\chi_z^2 = a_{..}t_z = \frac{a_{..}}{S_{xx}} \frac{(S_{xz})^2}{S_{zz}} \tag{12}$$

represents information using the (Z) system, the difference

$$\chi_{c-z}^2 = a_{..}(\theta - t_z) = \frac{a_{..}}{S_{xx}} \left(\sum \frac{d_j^2}{a_{.j}} - \frac{S_{xz}}{S_{zz}} \sum d_j z_j \right) \tag{13}$$

measures the information which was lost due to use of the (Z) system.

Since $a_{..}\theta$ requires $(m - 1)$ degrees of freedom and $a_{..}t_z$ only 1 degree of freedom, the lost information can be tested on the assumption that $a_{..}(\theta - t_z)$ is distributed as χ^2 with $m - 2$ degrees of freedom.

In example 1, we have

$$a_{..}\theta = 32.055 \quad (\text{D.F.} = 4)$$

$$\underline{a_{..}t_z = 28.082 \qquad (\text{D.F.} = 1)}$$

"lost information"   3.973   (D.F. = 3) $P = 0.28$

Hence, the assumed score system which was used by the authors did not cause a significant loss in information.

Assigning same score to adjacent categories.

The experimenter may start out with more observation categories than are biometrically significant. It may either be reflected in too small $a_{.j}$'s in some categories or in the fact that the computed score appears in a rank order that is biologically unsound — such as $c_j < c_{j+1} > c_{j+2}$ (where one would only accept $c_j < c_{j+1} < c_{j+2}$).

In these cases it is easy to form a combined score for adjacent categories.

$$c_{j,j+1,\dots,j+p} = \frac{d_j + d_{j+1} + \cdots + d_{j+p}}{a_{.j} + a_{.j+1} + \cdots + a_{.j+p}} \tag{14}$$

A new ratio, θ', is computed with the following substitution:

$$\theta' = \frac{1}{S_{xx}} \left(\frac{d^2}{a_{.1}} + \cdots + \frac{d_{j-1}^2}{a_{.j-1}} + \frac{(d_j + \cdots + d_{j+p})^2}{a_{.j} + \cdots + a_{j+p}} + \cdots + \frac{d_m^2}{a_m} \right) \tag{15}$$

and the information lost by the combination of adjacent scores is esti-

mated by the difference

$$\chi^2 = a_{..}(\theta - \theta') \tag{16}$$

with degrees of freedom equal to the number of categories that were lost by the combination, in the above case equal to p.

Pooling scores from several experiments.

Computation of appropriate scores of a certain type of bio-assay is only useful if one can arrive at a set of scores that is appropriate for all experiments of the same type, and if it can be shown that one such set will cause no significant loss of information to any single experiment of the type.

If several experiments are carried out with the same type of reagents and observations are made under the same condition, a common score system for all experiments can be computed in the following way.

Observations are grouped in the same categories (j), and doses are expressed in logarithms; if coded, the code interval means the same log interval in all experiments.

Let α_{ijg} be the number of individuals observed for the ith dose, in the jth category in the gth experiment. Then

$$d_{j.} = \sum_g d_{jg}$$

$$\text{and } a_{.j.} = \sum_i \sum_g a_{ijg}$$

$$\text{The system } c_{j.} = \frac{d_{j..}}{a_{.j.}}$$

will then be the system of scores that maximizes the variance term of the common slope in relation to the total variance corrected for individual means. The test for the general application of this common score system will consist in applying the χ^2-test for lost information. In computing the several items for this test it is useful to conduct the following computational checks:

$$\sum_g S(c_{j.})a_{.jg} = 0$$

$$\sum_g S(c_{j.})^2 a_{.jg} = \sum_g Sd_{jg}(c_{j.})$$

By computing the common score system ($m - 2$) degrees of freedom are used. Theoretically, this loss of d.f. would have to be distributed some way over all the experiments from which the score system is derived. If the number of experiments is large the individual χ^2-test will not be influenced by disregarding this difference.

Example 2.

Appropriate scores for bio-assay of tetanus toxoid in mice.

An inter-institutional study of the precision of tetanus toxoid assays in mice was reported a few years ago (5). This author who analyzed the data suggested a scoring system which was based on a graphical estimate that gave an apparent linear regression of mean score to log dose of toxoid. The set of 96 experiments made with various toxoids in six different laboratories provides an opportunity to apply the above method and test the validity of applying a uniform score system to all experiments.

The technique of the assay consisted in injecting graded doses of tetanus toxoid in groups of mice. Fourteen days later a measured dose of tetanus toxin was injected and deaths recorded daily for 7 days. On the seventh day survivors were recorded in two categories, survival with marked symptoms of tetanus, and survival without symptoms.

TABLE 3
Experiment I D-11. Tetanus toxoid assay.
Laboratory I. Toxoid *D*.

Dose of Toxoid	Coded log dose (x)	Number of mice dead on day						Survivors with tetanus	Survivors no sign	Total $a_{i.g}$
		2	3	4	5	6	7			
0.0125 ml	+1.5				1				5	6
0.00625 ml	+ .5	1	2					1	2	6
0.003125 ml	− .5	3	1					2		6
0.00156 ml	−1.5	4	1	1						6
$a_{.jg}$		8	4	1	1	0	0	3	7	24 ($a_{..g}$)
d_{jg}		−7.0	−1.0	−1.5	1.5	0	0	−0.5	8.5	0

$$S_{xx} = 30.00$$

Table 3 gives details of a protocol of one of the 6 experiments. There were five toxoids, one of which was assayed twice in each of 8 laboratory series. Each experiment was duplicated by random subdivision of each group of mice into two.

The designation of the experiments by ID11 means that it was part of laboratory I's series on toxoid *D*, first experimental day, first subgroup.

Table 3 also shows the value of the pertinent elements $a_{..}$, $a_{.j}$, S_{xx}

and d_j . These elements were summed over all 96 experiments and the resulting sums are shown in Table 4.

<div align="center">TABLE 4</div>

Computation of common score system for 96 individual experiments

Category	No. mice ($a_{.j}$)	$d_{j.}$	Score	
			$c_{j.} = d_{j.}/a_{.j.}$	$(d_{j.})^2/a_{.j.}$
Death 2 days	500	-277.9407	$-.5559$	154.59
Death 3 days	211	$- 64.7805$	$-.3070$	19.89
Death 4 days	45	$- 7.4286$	$-.1651$	1.23
Death 5 days	28	$+ 1.5000$	$+.0536$	0.08
Death 6 days	29	$+ 4.3787$	$+.1510$	0.66
Death 7 days	26	$- 9.4412$	$-.3631$	3.43
Survivors with tetanus	427	$+ 56.4701$	$+.1322$	7.47
Survivors, no signs	524	$+297.2422$	$+.5673$	168.61

<div align="center">Sums 1790 ($a_{...}$)   0.0000 $\theta = 355.87$</div>

$$S_{xx} = 1430.07$$
$$\chi^2 = a_{...}\theta = \frac{1790 \times 355.87}{1430.07} = 445.44 \ (\text{D.F.} = 7)$$

The appropriate scores are shown in the fourth column ($c_{j.}$). Comparison with the observation categories shows that the score is increasing with prolonged survival except in the categories for deaths on 5, 6 and 7 days. Since this sequence is biologically unacceptable, and the numbers in these categories are small, it is logical to test whether the irregularity of the computed scores is merely the result of random variation.

Hence, a combined score is computed for a category comprising 5-7 days (c_{5-7}).

$$c_{5-7} = \frac{1.5 + 4.3787 - 9.4412}{28 + 29 + 26} = -\frac{3.5625}{83} = -.0429$$

This score follows the biological rank of the chosen categories. A test whether the combination of the three categories causes significant loss of information is indicated by equations (15) and (16) above.

$$\chi^2 = \frac{1790}{1430.07}\left(0.08 + 0.66 + 3.43 - \frac{(3.5625)^2}{83}\right) = \frac{1790 \times 4.02}{1430.07}$$

$$= 5.03 \quad (\text{D.F.} = 2) \quad (P = 0.08)$$

Thus, there is a probability of 0.08 that the irregularity of the three categories is due to random variation, and we can accept the new combined score for all three categories.

The score system (Z) which the author suggested in previous publications was as follows:

		assumed score (z_i)
Death in less than	2 days	0
Death in	3 to 4 days	2
Death in	5 to 7 days	3
Survival with tetanus		4
Survival with no signs		6

Testing the fit of this assumed system we compute,

$$\sum a_{.j.}(z_i) = 5613$$
$$\sum a_{.j.}(z_i)^2 = 27467$$
$$\sum d_{j.}(z_i) = 1854.2279 = S_{xz}$$

and

$$S_{zz} = 27467 - (5613)^2/1790 = 9866.01$$

$$a_{..}t_z = \frac{1790}{1430.07} \cdot \frac{(1854.2279)^2}{9866.01} = 436.19$$

$$a_{..}(\theta - t_z) = 445.44 - 436.19 = 9.25 \qquad (\text{D.F.} = 6)$$

The fit of the assumed score for all 96 experiments together is quite acceptable $(P = 0.16)$. If the combined score for days 5-7 is used, the difference χ^2 is $9.25 - 5.03 = 4.22$ with 4 degrees of freedom. $(P = 0.39)$. Since the earlier report indicated that the mean slope for each set of experiments varies significantly between laboratories, there are reasons to test whether the overall score systems $(C.)$ and (Z) fit all laboratory sets without significant loss in information from that which would have been obtained by applying individual score systems (C_g) for each laboratory.

Table 5 gives these individual systems transformed linearly so that scores for the highest and lowest category are 6 and 0, respectively. The variation in score from laboratory to laboratory for the four "free" categories is quite impressive, but the chi-square value of the last columns indicate that the variation from the overall score system $(C.)$ indicated at the bottom of the table is not unreasonable. That of Laboratory V is the largest with $\chi^2 = 11.04$ $(P = 0.025)$.

The assumed score system shows in 5 of 8 laboratories a poorer fit with two chi-square values (Lab. IVF and V) having probabilities less than 1%.

TABLE 5

Overall computed (C.) and assumed (Z) score systems compared to systems (C_g) computed for each of eight laboratory sets.

Laboratory	Mice used ($a_{..g}$)	Score for category						x^2 (C_g) Indiv. System	Difference x^2 for system		D.F.
		Death				Survival			(C.)	(Z)	
		2	3	4	5–7 days	symp-toms	no sign				
I	240	0	2.07	1.40	2.68	3.21	6	86.38	1.24	0.99	4
II	215	0	2.82	6.22	5.40	6.18	6	22.43	4.36	3.39	4
III-1	216	0	1.38	−1.65	−0.90	2.23	6	46.70	5.52	6.91	4
III-2	214	0	1.51	2.90	−0.62	3.05	6	72.07	4.40	5.74	4
IV-M	213	0	0.64	−0.53	2.42	4.56	6	110.55	9.93	7.37	4
IV-F	216	0	0.26	1.46	1.00	2.08	6	87.26	6.37	15.02	4
V	263	0	0.62	5.37	6.18	3.98	6	94.54	11.04	13.77	4
VI	213	0	1.30	−0.07	4.18	4.37	6	34.03	2.61	2.70	4
Overall (C.)	1790	0	1.33	2.09	2.74	3.68	6	440.41	—	—	
(Z)	1790	0	2	2	3	4	6	436.19	4.22	—	4

TABLE 6

Values of the Mean Slope for Each Laboratory as Computed with the
Two Systems

Laboratory	Mean Slope*		Weight of Mean Slope	Intrinsic error		
	(C)	(Z)	S_{xx_g}**	(C)	(Z)	(C_g)
I	5.13	5.14	21.47	.403	.402	.399
II	2.79	2.81	12.86	.807	.785	.717
III-1	3.00	2.80	12.96	.505	.515	.466
III-2	5.57	5.50	12.86	.360	.366	.344
IV-M	6.42	6.42	12.77	.259	.253	.236
IV-F	5.77	5.53	12.96	.317	.346	.298
V	3.75	3.66	30.14	.496	.508	.452
VI	2.97	3.05	12.67	.586	.587	.559

*Increase in score per increase in log dose.
**x in log dose instead of coded values.

Table 6 gives the values of the mean slope for each laboratory as
computed with the two systems. The difference is very slight. The
last columns of Table 6 gives the effect on the precision of the assay
caused by use of three score systems, as found in variation of the
intrinsic error. This factor is usually computed as the square root of
the remainder variance (s) over the slope (b). The values in the table
have been computed with close approximation from the χ^2-values in
the following way.

$$(s/b)^2 = \left(\frac{1}{\chi^2} - \frac{1}{a_{..g}}\right)S_{xx_g} \tag{17}$$

which is derived from the approximation

$$s^2 = \frac{1}{a_{..g}}\left(S_{ss} - \frac{S_{xs}^2}{S_{xx}}\right) \tag{18}$$

and the equations

$$b = \frac{S_{xs}}{S_{xx}} \tag{19}$$

and

$$\chi_s^2 = \frac{a_{..g} \cdot (S_{xs})^2}{S_{xx} \cdot S_{ss}}$$

Relative efficiency

It may be of interest to express the loss due to the assumed score-system in terms of efficiency. Since the relative efficiency of two methods of conducting a bio-assay is expressed by the reciprocal ratio of the values of $(s/b)^2$ obtained in each instance, we have a measure of relative efficiency in the term

$$E = \left(\frac{s_1}{b_1}\right)^2 \cdot \left(\frac{b_2}{s_2}\right)^2 \tag{20}$$

E will express the relative efficiency of method (2) as a fraction of that of method (1). Using this expression for two score systems (C) and (Z) we obtain

$$E = \frac{\chi_z^2(a_{..g} - \chi_c^2)}{\chi_c^2(a_{..g} - \chi_z^2)} = \frac{t_z(1 - \theta)}{\theta(1 - t_z)} \tag{21}$$

For example, if the overall system $(C.)$ is applied to the data for Laboratory I instead of the individual system (Cg) for that Laboratory, we find

$$E = \frac{85.14(240 - 86.38)}{86.38(240 - 85.14)} = 0.978$$

TABLE 7

Relative efficiency of assay when overall score systems $(C.)$ and (Z) are applied instead of individual score system $(C_g.)$

Laboratory	Relative efficiency of	
	System $(C.)$	System (Z)
I	0.978	0.982
II	0.788	0.834
III-1	0.854	0.819
III-2	0.911	0.885
IV-M	0.830	0.871
IV-F	0.883	0.741*
V	0.829	0.790*
VI	0.910	0.907

*Significant loss in information ($P < 0.01$).

Table 7 presents the relative efficiency for each laboratory when the computed overall system $(C.)$ and the assumed system (Z) is applied, instead of individual laboratory systems. In these experiments, apparently 20 per cent efficiency loss is not to be considered significant.

Distribution of χ^2_{c-s}

Since the assumed system (Z) can be considered adequate for most experiments it is of interest to examine the distribution of the statistic χ^2_{c-s} for each individual small experiment. Assuredly, the assumption that it follows the χ^2-distribution is being stretched rather far, considering that only 18 animals were used in most of the 96 subgroups experiments.

TABLE 8

Distribution of χ^2_{c-s} for 93 experiments.
Compared to the probability of the theoretical χ^2-distribution.

P	Degrees of freedom				Total	
	1	2	3	4	observed	expected
.0 – .01					0	0.9
.01– .05		1			1	3.7
.05– .10			2		2	4.7
.10– .20		3	5	2	10	9.3
.20– .50	4	10	15	5	34	27.9
.50– .80	1	15	13	3	32	27.9
.80– .90		2	5		7	9.3
.90– .95	2	1			3	4.7
.95– .99		2	2		4	3.7
.99–1.00					0	0.9
	7	34	42	10	93	93.0

Table 8 presents the distribution of χ^2_{c-s} for the number of degrees of freedom which exist considering the categories in which observation were found. In three experiments, only two categories contained observations; hence any system will fit these experiments and there are no degrees of freedom.

The distribution tends toward smaller values of χ^2_{c-s} than expected for a theoretical distribution, but considering the small number of animals on which the computations are based, the agreement is quite satisfactory.

REFERENCES

1. J. Ipsen: Circulating antitoxin at the onset of diphtheria in 425 patients. *J. Immunology 54*, 325–347, 1946.

2. J. Ipsen and H.-I. Toft: Biological assay of strychnine in young mice using death time and convulsion time as biometrical scores. *Acta pharmacologica scand. 2*, 17, 1946.

3. J. Ipsen: The Effect of Environmental Temperature on the Immune Response of Mice to Tetanus Toxoid. *J. Immunology 69*, 273–283, 1952.

4. R. S. Ecke, A. G. Gilliam, J. C. Snyder, A. Yeomans, C. S. Zaraforetis, and E. S. Murray: The Effect of Cox-type Vaccine on Louse-borne Typhus Fever. *Amer. J. Trop. Med. 25*, 447–464, 1945.

5. J. Ipsen: Precision of Potency Assay of Alum-Precipitated Tetanus Toxoid in Mice. *J. Immunology 70*, 171–180, 1953.

MATRICES IN QUANTAL ANALYSIS

P. J. Claringbold

*Department of Veterinary Physiology, University of Sydney,
New South Wales, Australia*

INTRODUCTION

Fisher (1954) has recently discussed the various transformations of probability used in the analysis of binomial data, and in that paper a full account of the statistical theory is given. While the assumption is often made that a distribution of thresholds must be postulated before efficient analysis may be made of binomial data (Finney, 1952a), Fisher (1954) has clearly demonstrated that this is unnecessary. Transformations of the expected proportion responding may thus be simply regarded as a different scale for the measurement of response. In this paper practical methods of relating the binomial variable to the coordinates of experimental designs are given, and matrix methods are employed so that the results may immediately be applied to any experiment with known design matrix.

A considerable time has been devoted in the past to methods supposed to give quick estimation of parameters in quantal analysis. These graphical or semigraphical methods are usually employed in order to avoid efficient but tedious probit analysis in routine work. In this paper it will be shown that quick efficient solution is afforded by use of the angular transformation.

Recently Berkson (1953) has advanced a "simplified and quick" method for the estimation of parameters of binomial data by means of a modified logit technique. The method is still tedious when compared with the method exemplified here, which gives estimates equivalent to those derived by Berkson's method.

NOTATION

Scalars—letters in italics.
Vectors—small letters in heavy type.
Matrices—large letters in heavy type.
Transposition of matrices or vectors is indicated by a prime.
Unit matrix—**1**.
Diagonal matrices—**diag** $(\cdots)$, indicating diagonal elements.

An experimental design consists of N treatment combinations of k experimental variables or factors. If each of the factors is allocated a coordinate and scale of measurement the treatment combinations may be put in a one-one correspondence with row vectors of coordinate values, X_j, $j = 1, 2, \cdots, k$. Therefore N combinations may be represented in matrix notation by a *design matrix*, **D**, where,

$$\mathbf{D} = [X_j^i], \qquad i = 1, 2, \cdots, N.$$

Successive rows of this matrix give the coordinate values of the N combinations.

These combinations of treatments are administered to N groups of animals, so that in each group some, all or none respond. The sizes of the N groups may be summarized in a column vector **n**, the number responding in each group by the vector **a** and the observed proportion responding in each group in the vector **p**. The ith element in each of these column vectors corresponds to the set of treatment combinations specified in the ith row of the design matrix. Likewise the expected proportions responding may be summarized in the vector π, where $\pi = E(\mathbf{p})$.

The statistical problem considered is the relation of the expected proportion responding to the coordinates of the experimental design. Suppose the scale of measurement of this proportion undergoes a continuous transformation into another scale. Thus the ith value of π is a function of the ith value of some transformate, i.e.,

$$\pi_i = f(\rho_i). \tag{1}$$

The vector of N values of ρ_i is denoted ϱ. The Jacobian matrix of this transformation $\partial\pi/\partial\varrho$, is diagonal and denoted **diag.**(J).

The transformed vector may be related to the coordinates of the experimental design in a manner very similar to the standard regression analysis adopted with the normally distributed variable. The maximum likelihood procedure is outlined in the Appendix and is preferred since statistics estimated by this means have asymptotically normal distributions and the standard tests of significance may therefore be applied with some confidence. At the beginning of a regression analysis it is decided to relate expected response to certain functions of the coordinates of the experimental design. These may be powers, cross-products, certain sets of orthogonal functions and other functions. It is assumed that response may be expressed as a linear combination of these and a set of regression coefficients, $g' \leq N$ in number, $(g' = g + 1)$.

The coordinate functions may be evaluated at each point of the experimental design, that is for each row of the design matrix. The N by g' set of values may therefore be arranged as a matrix called the *matrix of coordinate functions.* It is assumed that the rank of this matrix is g', or that, in other words, the coordinate functions are linearly independent. Thus the ith value of the response transformate is a linear combination of a set of regression coefficients and the elements of the ith row of the matrix of coordinate functions.

$$\rho^i = x_0^i \beta_0 + x_1^i \beta_1 + \cdots + x_g^i \beta_g \tag{2}$$

or $\rho = \mathbf{X}\beta$, where β is a vector of regression coefficients. Since this transformation is linear it has the matrix $\mathbf{X}$ as its Jacobian, i.e. $\partial \rho^i / \partial \beta_s = x_s^i$, where x_s^i is the i, sth element of $\mathbf{X}$, the matrix of coordinate functions. In equation (2) small letters x are used to distinguish the values of the coodinate functions from those of the original coordinates X_i. In general the sth coordinate function x_s may be a function of any, some or all of the coordinates. One function is always defined, $x_0 = 1$, and is termed the identity. Estimates of β_0 give the response intercept when all other coordinate functions are zero.

Provided that it is possible to describe the response in terms of the functions chosen the solution is readily obtained in practice. Otherwise the iterative procedure used may take many steps and the regression coefficients estimated are biassed. This feature of regression analysis is fully discussed by Box and Wilson (1951) where the General Theory of Aliases is given in matrix notation.

<div align="center">EXAMPLES</div>

Examples involving a small number of experimental points must be chosen if they are to go on a page of the journal. The method of analysis and the laying out of computations in larger examples are identical in form to those illustrated, simply occupying more space. In the practical carrying out of certain matrix multiplications the computor needs to learn one new operation, namely, row into row or column into column multiplication.

Examples with diagonal information matrix.

Into this class fall all factorial and other orthogonal designs where the angular transformation is employed with equal group sizes. In certain other designs the information is readily reducible to the diagonal form by means of suitable coordinate transformations, for example the parallelogram designs of Claringbold, Biggers and Emmens (1953)

1. *Four point assay.*

Berkson (1953) advocates a "short cut minimum logit χ^2 method" for the four point assay. The analysis takes a full page of the journal in small type and requires reference to tables and nomographs not freely available. In Berkson's example the ratio of successive doses of the standard preparation (S) and the unknown preparation (U) was 1.5.

$$
\begin{array}{cc}
1/nw & (\mathbf{X'X})^{-1}
\end{array}
\qquad \mathbf{X'} = [x_i^s]
$$

					S	U			$\check{\beta}$	$\beta_{(1)}$	s_β
$820.7/24$	0.25	$\cdot$	$\cdot$	1	1	1	1	β_0	47.0	46.83	2.93
		0.25	$\cdot$	-1	-1	1	1	β_1	4.5	4.33	2.93
	$\cdot$	$\cdot$	0.25	-1	1	-1	1	β_2	15.5	15.43	2.93

$$
\begin{array}{llllll}
& \mathbf{p'} & 25 & 67 & 29 & 88 \qquad \text{Log } M = -0.2906 \\
\chi^2_{(1)} = 1.06, & \mathbf{r'} & 30 & 55 & 33 & 70 \qquad\qquad \pm 0.1973 \\
P > 0.05 & \mathbf{\varrho'}_{(1)} & 27 & 58 & 36 & 67 \qquad M = 0.893 \\
& \mathbf{\psi'}_{(1)} & 30.1 & 54.9 & 32.7 & 69.6 \qquad \cdot \{0.760 - 1.039\}
\end{array}
$$

The design matrix may be seen in the second and third rows of $\mathbf{X'}$. The regression coefficient β_0 estimates the mean response, β_1 the difference between samples and β_2 the slope of the dose response lines. The computational operations comprise:

(i) Form $\mathbf{X'}$: At most four regression coefficients are required to describe the experimental data. Three have been estimated above while the fourth could be a measure of the departure from parallelism between the two dose response lines.

(ii) Form $(\mathbf{X'X})^{-1}$: Since the coordinate functions are orthogonal this matrix is diagonal. The orthogonality of the functions may be checked by seeing that the sum of the cross products of equivalent elements in different rows is zero. The diagonal elements of $(\mathbf{X'X})^{-1}$ are the reciprocals of the sums of squares of the individual rows of $\mathbf{X'}$. The scalar $1/nw$ may immediately be tabulated, and together with $(\mathbf{X'X})^{-1}$ gives the variance-covariance matrix of the regression coefficients. Thus the standard errors of these coefficients may be immediately determined and tabulated. For example,

$$
s(\beta_0) = \left\{ \frac{820.7}{24} \times 0.25 \right\}^{1/2}
$$

(iii) Write the observed proportions responding under the appropriate column of $\mathbf{X}'$, i.e. complete the row labelled $\mathbf{p}'$.

(iv) Use Fisher and Yates (1953) Table XII or Finney (1952b) Appendix Table XI to tabulate the empirical angles as a row $\mathbf{r}'$.

(v) Calculate provisional estimates of the regression coefficients by row into row multiplication, thus

$$x_1^s r_1 + x_2^s r_2 + x_3^s r_3 + x_4^s r_4, \qquad s = 0, 1, 2.$$

and multiply by the sth element of $(\mathbf{X}'\mathbf{X})^{-1}$ to give β^s. For example $47.0 = (1 \times 30 + 1 \times 55 + 1 \times 33 + 1 \times 70) \times 0.25$

(vi) Form the row of expected responses by column into column multiplication, thus

$$\rho^i = x_i^1 \beta^1 + x_i^2 \beta^2 + x_i^3 \beta^3 + x_i^4 \beta^4, \qquad i = 1, 2, 3 \text{ or } 4.$$

For example, $27 = (1 \times 47.0) - (1 \times 4.5) - (1 \times 15.5)$

(vii) Use Finney (1952b) Appendix Table XIII to tabulate, or Fisher and Yates (1953) Table XIV to calculate the row of working angles, ψ'.

(viii) Repeat operation (v) using the working angles. The iterative procedure may be considered complete if the 10% criterion is adopted (see below). Otherwise operations (vi) and (vii) are alternated, building up new rows of working angles and new columns of estimates.

(ix) Log relative potency, $\log M$, is given by the ratio β_1/β_2, and a negative sign is taken since the response to U is less than to S.

(x) The variance of $\log M$, $V(\log M)$, is given by Fieller's formula, discussed by Finney (1952b). Since β_1 and β_2 are independent, in this example, the formula simplifies,

$$V(\log M) = 1/\beta_2^2 \cdot [V(\beta_1) + (\log M)^2 V(\beta_2)]$$

$$= 820.7/(96\beta_2^2) \cdot [1 + (\log M)^2]$$

(xi) Fiducial limits are obtained in the log scale and then transformed into the arithmetic scale by taking antilogarithms to the base 1.5.

(xii) The χ^2 goodness-of-fit is formed from the difference of the weighted sum of squares of the working angles and the weighted sum of cross

products between the working angles and the expected response. It has $4 - 3 = 1$ degree of freedom. Thus

$$\chi^2 = [(30.1^2 + 54.9^2 + \cdots) - (30.1 \times 27.1 + 54.9$$

$$\times 57.9 + \cdots)] \cdot 24/820.7 = 1.06$$

The calculation of relative potency is unnecessary in this example if all that is required is the answer to the question, "Does the activity of U differ significantly from that of S?" The significance of β_2 tests this hypothesis,

i.e. $t_{(\infty)} = 1.48$, $P > 0.05$

Berkson's estimate of M for this example is 0.887 {0.746 − 1.056}, which differs by only 0.006 from ours. This is a negligible difference in terms of the confidence interval. Berkson (1953) mistakenly uses the antilogarithm of the standard error of log relative potency as standard error of relative potency. The fiducial limits are the appropriate measure of accuracy here. It may be noted that the angular transformation gave a narrower fiducial interval than logits.

2. Parabola design.

Examples of parallelogram designs have been given by Claringbold, Biggers and Emmens (1953). In that paper it was suggested that the principle developed could be extended. The present example is extracted from a 2×5^2 factorial, for illustrative purposes, and has the design matrix contracted into groups of three rows:—

$$
\mathbf{D} =
\begin{bmatrix}
-1 & 0 & 1 & 2 \\
0 & -1 & 0 & 1 \\
1 & 0 & 0 & 2
\end{bmatrix}
\begin{matrix}
X_1 & & X_2 & \\
\end{matrix}
$$

Thus at the zero level of X_1, the levels of $\log_2$ dose, namely X_2, were lowered by one unit. The design presupposes a quadratic relationship between response and X_1, it being known in advance that the middle level of X_1 would markedly increase response to dose.

It was decided to relate response to four coordinate functions, $x_0 = 1$, $x_1 = X_1$, $x_2 = (X_1)^2$ and $x_3 = X_2$. The transposed matrix of coordinate functions is,

$$\mathbf{X'}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & -1 & 0 & 1 & 0 & 1 & 2 \end{bmatrix} \begin{matrix} x_0 \\ x_1 \\ x_2 \\ x_3 \; . \end{matrix}$$

Formation of the sums of squares and cross products between pairs of rows of this matrix show these functions to be non-orthogonal. The variance covariance matrix is therefore non-diagonal.

Premultiplication of the matrix $\mathbf{X'}$ by the matrix $\mathbf{C'}$ where,

$$\mathbf{C'} = \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot \\ -2 & \cdot & 3 & \cdot \\ \cdot & \cdot & -1 & 1 \end{bmatrix}$$

gives the transformed matrix of coordinate functions, $\mathbf{X^{*'}}$ tabulated below. The computational procedures are similar to those given above for the four point assay except for the stages after iteration when the unstarred coordinate functions are reintroduced.

In this analysis the 100% values are initially given the value 84 in conformity with the correction discussed by Claringbold, Biggers and Emmens (1953) and used with the empirical angular transformation. The χ^2 goodness of fit is not significantly enlarged and it may therefore be assumed that the regression model adequately described the data, i.e.

$$\rho = 46.1 - 2.9x_1^* + 3.0x_2^* + 32.1x_3^*$$

The vector of estimated regression coefficients may be transformed so that they relate response to the original unstarred coordinate functions, thus:

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 1 & \cdot & -2 & \cdot \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 3 & -1 \\ \cdot & \cdot & \cdot & 1 \end{bmatrix} \cdot \begin{bmatrix} 46.1 \\ -2.9 \\ 3.0 \\ 32.1 \end{bmatrix} = \begin{bmatrix} 40.1 \\ -2.9 \\ -23.1 \\ 32.1 \end{bmatrix}$$

or $\beta = C\beta^*$, see Appendix equation 8.

$$\frac{1}{nw}\,(X^{*\prime}X^{*})^{-1} = \frac{820.7}{24}\begin{bmatrix} 0.1111 & \cdot & \cdot & \cdot \\ \cdot & 0.1667 & \cdot & \cdot \\ \cdot & \cdot & 0.0556 & \cdot \\ \cdot & \cdot & \cdot & 0.1667 \end{bmatrix}$$

$$X^{*\prime} = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & -2 & -2 & -2 & 1 & 1 & 1 & 1 \\
-1 & 1 & -1 & 0 & 1 & 0 & -1 & 0 & 1
\end{bmatrix}$$

	$\hat{\beta}^*$	$\beta^*_{(1)}$	$\beta^*_{(2)}$	$\beta^*_{(3)}$	s^*_β	
	45.3	46.1	46.3	46.1	2.0	x_0^*
	-2.3	-2.7	-2.9	-2.9	2.4	x_1^*
	3.3	3.1	2.9	3.0	1.4	x_2^*
	32.3	32.2	31.9	32.1	2.4	x_3^*

p'	8	63	100	8	25	88	4	46	100
r'	16	53	84	16	30	70	12	43	84
$\varrho'_{(1)}$	19	51	83	6	39	71	14	46	79
$\psi'_{(1)}$	16.6	52.5	86.5	19.5	30.5	69.7	11.8	42.8	84.6
$\varrho'_{(2)}$	19.7	51.9	84.1	7.7	39.9	72.1	14.3	46.5	78.7
$\psi'_{(2)}$	16.7	52.6	87.1	21.1	30.6	69.6	11.8	42.7	84.4
$\varrho'_{(3)}$	20.2	52.1	84.0	8.6	40.5	72.4	14.4	46.3	78.2
$\psi'_{(3)}$	16.7	52.6	87.0	19.8	30.6	69.6	11.8	42.8	84.2

$$\chi^2_{[15]} = 8.1, \quad 0.3 > P > 0.2$$

Thus
$$\rho \doteq 40.1 - 2.9x_1 - 23.1x_2 + 32.1x_3$$
$$= 40.1 - 2.9X_1 - 23.1(X_1)^2 + 32.1X_2$$

Since the starred regression coefficients are independent the variances of any linear combinations of them are simply determined. For example since $\beta_0 = \beta_0^* - 2\beta_2^*$, then $V(\beta_0) = V(\beta_0^*) + 4V(\beta_2^*)$.

Example with non-diagonal information matrix.

The type of analysis demonstrated below must be used with all non-orthogonal and/or badly designed experiments or when group sizes are unequal. The process must also be used with all transformations other than angular with the added disadvantage that the variance-covariance matrix changes, albeit slowly, at each stage of iteration. The example chosen is a 2^3 factorial experiment with unequal group sizes. One factor is replication so that the parameters requiring estimation are the identity, replicate difference, the effect of each treatment and the interaction between treatments i.e. five regression coefficients. The coordinate functions are $x_0 = 1$, $x_R = X_R$, $x_1, = X_1$, $x_2 = X_2$, $x_3 = X_1X_2$.

The computational procedures comprise:

(i) Determine a working unit of weight. In this example the average group size was given unit weight.

(ii) Write down the matrix coordinate functions, $\mathbf{X}'$.

(iii) Multiply successive columns of this matrix by the appropriate group weight. This is tabulated unchanged in the first row of the weighted matrix of coordinate functions, $\mathbf{X}'\mathbf{W}$.

(iv) Form the matrix $\mathbf{X}'\mathbf{W}\mathbf{X}$ by row into row multiplication of $\mathbf{X}'$ into $\mathbf{X}'\mathbf{W}$. Thus the second row, third column element of $\mathbf{X}'\mathbf{W}\mathbf{X}$ is obtained by the row 2 into row 3 multiplication, viz.

$$(-1 \times -1.08) + (-1 \times -1.08) + (-1 \times 0.92) + (-1 \times 1.03)$$
$$+ (1 \times -1.08) + (1 \times -1.08) + (1 \times 0.70) + (1 \times 1.03) = -0.22$$

(v) Form the inverse of this matrix $\mathbf{X}'\mathbf{W}\mathbf{X}$ using the method of say Fox (1950) and Fox and Hayes (1951). This involves formation of an upper triangular matrix $(\mathbf{A})$ and then the inverse. The full layout is shown, excluding check lines, but will not be explained as this is fully carried out by Fox and Hayes (1951). The inverse matrix is the

Ratio no. responding to no. per group

Design matrix

Factor A X_1	Factor B X_2	First Replicate $X_R = -1$	Second Replicate $X_R = 1$
-1	-1	3:20	4:20
-1	1	15:20	15:20
1	-1	4:17	5:13
1	1	17:19	17:19

Average group size = 18.50

$\mathbf{X'}$

x_0	1	1	1	1	1	1	1	1
x_R	-1	-1	-1	-1	1	1	1	1
x_1	-1	-1	1	1	-1	-1	1	1
x_2	-1	1	-1	1	-1	1	-1	1
x_3	1	-1	-1	1	1	-1	-1	1

$\mathbf{X'W}$

x_0	1.08	1.08	0.92	1.03	1.08	1.08	0.70	1.03
x_R	-1.08	-1.08	-0.92	-1.03	1.08	1.08	0.70	1.03
x_1	-1.08	-1.08	0.92	1.03	-1.08	-1.08	0.70	1.03
x_2	-1.08	1.08	-0.92	1.03	-1.08	1.08	-0.70	1.03
x_3	1.08	-1.08	-0.92	1.03	1.08	-1.08	-0.70	1.03

X'WX

$$\begin{bmatrix} 8.00 & -0.22 & -0.65 & 0.43 & 0.43 \\ -0.22 & 8.00 & -0.22 & 0.22 & 0.22 \\ -0.65 & -0.22 & 8.00 & 0.43 & 0.43 \\ 0.43 & 0.22 & 0.43 & 8.00 & -0.65 \\ 0.43 & 0.22 & 0.43 & -0.65 & 8.00 \end{bmatrix}$$

A

$$\begin{bmatrix} 2.83 & -0.08 & -0.32 & 0.15 & 0.15 \\ 0 & 2.83 & -0.08 & 0.08 & 0.08 \\ 0 & 0 & 2.82 & 0.17 & 0.17 \\ 0 & 0 & 0 & 2.82 & -0.25 \\ 0 & 0 & 0 & 0 & 2.81 \end{bmatrix}$$

$1000(\mathbf{X'WX})^{-1}$

$$\begin{bmatrix} 127 & 4 & 12 & -8 & -8 \\ 4 & 126 & 4 & -4 & -4 \\ 12 & 4 & 127 & -8 & -8 \\ -8 & -4 & -8 & 127 & 12 \\ -8 & -4 & -8 & 12 & 127 \end{bmatrix}$$

	$\check{\beta}_w$	$\beta_{(1)}$	$\beta_{(2)}$	S_β
x_0	47.5	47.4	2.4	
x_R	1.5	1.4	2.4	
x_1	5.0	5.0	2.4	
x_2	18.5	18.4	2.4	
x_3	1.2	1.1	2.4	

1000T

$$\begin{bmatrix} 120 & 120 & 138 & 121 & 130 & 130 & 112 & 129 \\ -136 & -136 & -100 & -129 & 136 & 136 & 100 & 129 \\ -129 & -130 & 138 & 121 & -120 & -121 & 112 & 129 \\ -118 & 128 & -141 & 131 & -128 & 118 & -113 & 123 \\ 127 & -118 & -140 & 131 & 118 & -128 & -113 & 123 \end{bmatrix}$$

$\mathbf{p'}$	15	75	24	90	20	75	38	90
$\boldsymbol{\gamma'}$	23	60	29	72	27	60	38	72
$\boldsymbol{\theta}'_{(1)}$	24	58	31	71	27	61	34	72
$\boldsymbol{\psi}'_{(1)}$	22.8	60.0	29.4	71.6	26.6	60.0	38.2	71.4
$\boldsymbol{\psi}'_{(1)w}$	24.6	64.9	27.0	73.5	28.8	64.9	26.9	73.3

$\chi^2_{(3)} = 2.11, \quad 0.7 > P > 0.5$

variance covariance matrix when multiplied by a scalar 820.7 ÷ 18.5. i.e. the reciprocal of the unit weight.

(vi) Form the transformation matrix (T) by column into column multiplication of the inverse matrix (excluding the scalar factor since this cancels out) into the weighted matrix of coordinate functions, $X'W$. Thus the first row, second column element of T is given by:

$(0.127 \times 1.08) + (0.004 \times -1.08) + (0.012 \times -1.08)$
$$+ (-0.008 \times 1.08) + (-0.008 \times -1.08) = 0.120.$$

In all the above matrix operations two additional decimal places were held at all stages. The figures have been rounded to save space. The sum of the elements of the first row of T should be unity and the remainder zero.

(vii) Write down row of observed percentage responding under each appropriate column of T.

(viii) Begin the iterative procedure in an exactly analogous manner to the above examples, obtain regression coefficients by row into row multiplication of T into the working of empirical angles, and new rows of expectation by column into column multiplication of regression coefficients into the matrix of coordinate functions.

(ix) The calculation of χ^2 goodness-of-fit requires an additional step at the end of iteration. A row is formed the elements of which are the product of the final working angles and the corresponding weight (first line of $X'W$). The sum of cross products of the elements of this and the previous line, minus the sum of cross products between the elements of this line and the elements of the line of expected responses above, multiplied by the value of the unit weight, gives a χ^2 goodness-of-fit test. Thus

$$\chi^2 = (24.6 \times 22.8 + \cdots) - (24.6 \times 24 + \cdots) \times \frac{18.5}{820.7}$$

$$= 2.11$$

DISCUSSION

The angular transformation may be used in two distinct types of experimental problem where the response variable is binomial. The first type of problem is exemplified by studies of some dose response law. If it is of interest to show a log normal tolerance distribution studies must be made with the probit transformation. If such a law has been established then the angular transformation may only be used as a very good approximation in the interval of 5–95% expected response. This has been practically demonstrated by Biggers (1951) and may be

observed by graphical comparison of the transformations (Finney, 1952b). Similar considerations apply to logits.

In the second type of problem such a response law may be unknown or not the prime interest of study. For example Cochran (1938) discusses field experiments with percentage data and does not mention the tolerance distribution. More recently Campbell, Hancock and Rothschild (1953) have applied the angular transformation to percentage dead or alive spermatozoa using the angular transformation simply as a convenient scale. Finally Fisher (1954) does not mention tolerance distributions but regards transformations simply as transformations of probability of response. In this type of problem the angular transformation is not an approximation but is a more convenient measure of response.

Various binomial transformations are used in the case where variation is quantal but in excess of the Bernoulli binomial distribution. The problem has been recently discussed by Bartlett (1954) and Anscombe (1954). In this case the distribution is unknown and the maximum likelihood procedure is inapplicable. While the angular transformation may be used here any procedures are largely arbitrary. Bartlett (1954) states that there are very few examples of binomial data in the literature which are not heterogeneous. In this laboratory it has been found that with strict randomisation of animals to experimental units and with stratified randomisation of operators to the experimental units the conditions for obtaining homogeneous results occur. At present thirty factorial or other complex experiments have been carried out with the binomial variable, analysed with the angular transformation and all but one found homogeneous. These are available on request, and have been published in a number of biological journals (Biggers & Claringbold, 1954; Claringbold, 1953). It seems that strict randomisation and control is essential.

The first two examples illustrate the simplicity of the estimation procedure advocated in this paper. Berkson (1953) makes an important point when he criticises the routine users of probit analysis who employ only one cycle of the iterative procedure following graphical estimation. The procedure leads to widely discrepant results in small samples, say of 50 animals. The time taken in routine analysis with the probit transformation is excessive and a frequent cause of this error, while in this type of work use of the angular transformation with equal steps in the log dose scale results in quick analysis. Groups of animals with expected responses outside the interval 5–95% may be ignored to preserve linearity.

When should iteration be stopped? One arbitrary but stringent

criterion suggested by Fisher and Yates (1953) is the acceptance of estimates which differ from those of the previous iteration, or the previous graphical estimates in the case of the first iteration, by less than 10% of their standard error. It is difficult to imagine any experimental situation where differences of this order could be considered important, after all, a statistic with infinite degrees of freedom must exceed zero by 196% of its standard error before being considered significant.

Large regions of 0 and 100% responses in factorial experiments have been discussed by Claringbold, Biggers and Emmens (1953). It was concluded that the best method was to use special experimental designs based on small scale pilot experiments. The second example is a simplified form of such a special design. The aim of the design was to obtain a precise estimate of slope with three levels of an additional variable. The doses were therefore spaced as widely as possible in the interval 5–95%. Another aim of this design was to establish more precisely the quadratic relationship of response to X_1 .

Alternatives to this design have disadvantages. The experiment could be designed as a 3^2 factorial with the dose interval halved so the responses avoided the extremes of the scale. This results in a loss of 75% of the information about slope. If a standard factorial were used without the reduction in scale interval there would be, depending on the centring, two groups with expectations very near 100% on the middle line or two groups with expectations near 0% on the outer lines. Linearity in angles would therefore be lost and to preserve linearity, analysis in terms of probits or logits would have to be made with resultant loss in simplicity.

In probit and logit analysis of factorial or other experiments where there are large regions of 0 and 100% response an additional difficulty arises. When the expected response approaches 0 and 100% the weight of a probit or logit approaches zero, and may reach zero in terms of the number of figures carried by the computer. If the number of experimental groups giving reasonable amounts of information is less than the number of regression coefficients requiring estimation, attempts to fit this number of regression coefficients are doomed to inaccuracy and near aliases. This is evinced by the misbehaviour of the numerical computation when logits or probits are used in such cases. The mathematical reason underlying this situation is that the rank of the information matrix becomes less than g', and before it may be inverted the linearly related coordinate functions must be eliminated. It can be seen therefore that problems of regions of 0 and 100% responses are not

confined to the angular analysis and are best avoided by designing the large scale experiment on the basis of small pilot tests.

When the expectation of every group is at one end of the response scale the number of animals responding (or alternatively, those not responding, whichever is the smaller may be regarded as a Poissonian variable and the square root transformation used. This transformation has the same effect as the angular in equalising information.

With the third example the awkwardness of non-orthogonal or unequally weighted data is illustrated. This whole process must theoretically be repeated at each cycle for transformations other than the angular, since the information changes with each cycle. Usually it is sufficient to calculate information initially from the trial solution, run through the iterative procedure to convergence and then recompute information and check convergence.

While the first example took 11 minutes and the second 33 minutes to analyse, including checks, the third involved some 126 minutes of computational time, most of which was spent in forming and inverting the information matrix. If this had been repeated for different transformations the information matrix would need to be formed at least twice thus involving a tedious 4 or more hours work. This is a small factorial experiment. Consider the analysis of a 2^6 experiment in say probits. If we wish to estimate a mean, all main effects and first order interactions, 22 parameters require estimation. It would appear that the inversion of the 22 by 22 information matrix renders probit analysis prohibitive.

Finney (1952a) in the discussion of a 2^3 factorial experiment used by Potter and Gillham (1946) suggests a different approach. In this experiment a dose response line was determined using five doses under the eight sets of conditions specified in the 2^3 design. The method of analysis was: (1) compute the eight dose response lines using probits, (2) test for parallelism of the lines and their homogeneity, and estimate a common slope, (3) compute the weighted means ($\bar{x}$ and $\bar{y}$) of each dose response line, (4) perform an *unweighted analysis* of these means to obtain estimates of the eight parameters required to describe each of these, i.e. grand mean, three main effects, three first order interactions and a second order interaction. All but the first of these parameters is a difference between four weighted means and the other four. (5) Divide these seven differences of the $\bar{y}$ by the common slope and subtract from the corresponding difference of the $\bar{x}$, to give relative potency figures to which may be attached a standard error. This analysis is still tedious since eight lines must be fitted and iterated one at a time.

The most serious objection is the unweighted analysis of the weighted means. Their weights ranged from 24.2 to 115 and so the efficient analysis must be a weighted analysis. The approach runs into further complications when slope is affected by the treatments (see for e.g. Biggers, Claringbold and Emmens, 1954), when a separate weighted analysis of slope must be carried out. Unfortunately this example is heterogeneous and many 100% responses are present, thus rendering it inappropriate for angular analysis.

It may be concluded that only with the angular transformation are the modern experimental designs to be applied freely with the binomial variable. It cannot however, be concluded that this transformation will be the best or appropriate in all experimental situations; it may only be said that it appears by far the simplest solution in laboratory experiments where rigid control and randomisation ensure the classical Bernoulli binomial distribution. Some have said that use of the angular transformation may give misleading results. The final arbiter on this point is the goodness-of-fit test. If this is satisfactory our assumptions of both transformation and form of the regression equation have not been disproved.

<div align="center">APPENDIX</div>

Fisher (1954) has given a full account of the binomial probability distribution, most transformations commonly used, the theory of the maximum likelihood procedure and its application to the present problem. The matrix equivalent equations are developed briefly here in order to justify the estimation procedure laid out in the examples.

Likelihood equations and information

The loglikelihood (Fisher, 1925, 1934) of the vector **a** defined above is given by,

$$L = \mathbf{a}'[\log \pi_i] + (\mathbf{n}' - \mathbf{a}')[\log (1 - \pi_i)] \tag{3}$$

The quantities in square brackets are column vectors of length N, the ith element of which is shown.

Thus loglikelihood is a function of the observations **a** and the expected proportion π apart from a constant which has been ignored. The expected proportion is a function of the transform (ϱ), which in turn is a function of the vector of regression coefficients, β.

The maximum likelihood estimate is obtained by the solution of the partial differential equations, $\partial L/\partial \beta = 0$. In terms of the transformations (1) and (2), as a column vector,

$$\partial L/\partial \beta = \mathbf{X}' \operatorname{diag}(J)\, \partial L/\partial \pi = 0 \tag{4}$$

The information in the sample with respect to the parameters is given by the negative of the expectation of the second partial differentials of the loglikelihood equation with respect to the parameters.

Thus $\mathbf{I} = \mathbf{X'WX}$ where $\mathbf{W}$ is the diagonal weight matrix of the transformation, defined,

$$\mathbf{W} = \{\mathrm{diag}(J)\}^2 \cdot \mathrm{diag}\{n_i/\pi_i(1 - \pi_i)\} \tag{5}$$

The variance covariance matrix of the regression coefficients is given by,

$$\mathbf{V} = \mathbf{I}^{-1}, \qquad \text{provided } [\mathbf{I}] \neq 0.$$

Solution of the estimation equations

The solution is based on the scoring system discussed by Fisher (1946) and described in the general case by Rao (1952). The vector of first partial differentials $\partial L/\partial \beta$ is expanded about a trial value indicated $\partial L/\partial \beta_{(0)}$ to the first order. In this equation the information matrix may replace the second partial differentials so that,

$$\partial L/\partial \beta = \partial L/\partial \beta_{(0)} + \mathbf{I}_{(0)} \Delta \beta_{(0)} = 0 \tag{6}$$

where $\Delta \beta$ is a vector of additive corrections. The bracketed subscripts denote evaluation at a trial value.

Solution is facilitated by the introduction of a *working vector variate* which may be defined in a number of alternative ways. For example,

$$\psi = \{\varrho - \mathrm{diag}(J)^{-1}\pi\} + \mathrm{diag}(J)^{-1}\mathbf{p},$$

where the quantity in braces is called the *minimum working variate* and the individual values of $\mathrm{diag}(J)^{-1}$ the *range*.

Substitution in equation (6) with considerable rearrangements gives, at the ith stage of iteration,

$$\beta_{(i+1)} = \beta_{(i)} + \Delta \beta_{(i)} = (\mathbf{X'W}_{(i)}\mathbf{X})^{-1}\mathbf{X'W}_{(i)}\psi_{(i)}$$

$$= \mathbf{T}_{(i)}\psi_{(i)}, \tag{7}$$

where $\mathbf{T}$ is termed the *transformation matrix*.

Linear transformations of the matrix of coordinate functions.

Suppose in equation (2) a linear transformation of rank g' is made on the matrix of coordinate functions by the square matrix $\mathbf{C}$. The inverse transformation is made on the vector of regression coefficients, i.e.

$$\varrho = \mathbf{XCC}^{-1}\beta = \mathbf{X}^*\beta^* \tag{8}$$

The maximum likelihood estimate of the transformed vector may be reached in the standard manner as outlined above. On reaching the solution the inverse transformation may simply be made from the relation $\beta = C\beta^*$. Thus the matrix C^{-1} need never be computed. This procedure will not affect estimates of regression coefficients as is shown,

$$\beta = C\beta^* = C(X^{*\prime}WX^*)^{-1}X^{*\prime}W\psi$$
$$= C(C'X'WXC)^{-1}C'X'W\psi = CC^{-1}(X'WX)^{-1}C'^{-1}C'\psi$$
$$= \text{equation 7}.$$

While it is always possible to find such a transformation which will render the information matrix diagonal it is only of use in practice when determined quickly. This is possible with a scalar weight matrix and where parallelogram designs and their extensions have been used.

The angular transformation

Fisher (1922) used the angular transformation in the study of binomial data. It is defined,

$$\pi_i = \sin^2 \rho_i$$

With this transformation the weight matrix becomes independent of the expected response. The matrix may therefore be computed once and for all at the beginning of the iterative procedure as it does not change from cycle to cycle. Standard errors of regression coefficients are therefore known in advance, and the design and allocation of experimental animals to groups may be adjusted to give any predetermined degree of accuracy. If the goodness-of-fit is unsatisfactory the assumptions on which these standard errors are based have been shown false and the observed variation must be used instead of the theoretical variation to judge the significance of regression coefficients. Their variances are enlarged by a heterogeneity factor, the goodness-of-fit χ^2 divided by its degrees of freedom (Finney, 1952a), and the t test of their significance is based on this number of degrees of freedom.

The weight matrix becomes,

$$\mathbf{W} = w \, \mathbf{diag}(n_i),$$

where $w = 1/820.7$ if ρ measured in degrees.

If the elements n_i are constant, i.e. if the group size is constant, say n, the weight matrix is given by,

$$\mathbf{W} = nw\mathbf{1}, \text{ a scalar matrix.}$$

If the weight matrix is scalar and the coordinate functions are orthogonal the information matrix is diagonal. The functions are orthogonal if,

$$\mathbf{X'X} = \mathbf{diag}(\cdots)$$

Granted a scalar weight matrix it may be cancelled from the estimation equations so the transformation matrix becomes,

$$\mathbf{T} = (\mathbf{X'X})^{-1}\mathbf{X'}$$

Claringbold, Biggers and Emmens (1953) showed that in factorial experiments where the total number of experimental animals exceeded about 300, the equation,

$\breve{\beta} = \mathbf{Tr}$, where $p_i = \sin^2 r_i$, and p_i is the lth observed proportion,

gave estimates of a very similar value to the maximum likelihood estimates. Also it was suggested that in small samples the provisional estimates should be obtained with this equation.

Unequal group sizes.

With unequal group sizes the analysis in fully efficient form must be weighted. In this case the information matrix is non-diagonal and it must be inverted using more complicated procedures than forming reciprocals of the individual elements, as is done in the case of a diagonal matrix.

An alternative approximate procedure is available when the experimental groups suffer small losses during the course of an experiment and when the variance of this loss is known. Fisher (1925) has discussed an analogous problem in the derivation of a combined estimate from a number of estimates with faulty weights. It was concluded that if the variance of the false weights about their mean value was small, the combined estimate suffered little loss in efficiency if the estimates were weighted in terms of the average weight. In the present problem a similar solution is offered by the use of the average group size $\bar{n}$ to form the weight matrix. If the loss of experimental material is a random binomial variable with expectation about 1–2% it may be shown that the use of the average group size results in little additional loss of information. The loss is NQw where Q is the probability of loss of an experimental animal. In the type of experiments carried out in this laboratory a loss of 1% of experimental animals is unusually large.

Analysis of variance.

When the estimates of the regression coefficients have been obtained the standard analysis of variance may be carried out if the information matrix is diagonal. The total sum of squares of the working variate is

partitioned into a series of independent sums of squares corresponding with treatment effects.

Goodness-of-fit.

The weighted deviations from regression give a χ^2 test (Fisher, 1954) with $(N - g')$ degrees of freedom.

$$\text{i.e.} \quad \chi^2 = \psi'\mathbf{W}\psi - \{\beta'\mathbf{I}\beta \text{ or } \psi'\mathbf{W}\varrho\}$$

Special case when $g' = N$.

Bailey (1951) has shown that when the number of regression co-efficients requiring estimation is equal to the number of experimental points the observed values may replace the expected in the estimation equations. In this case the working variate degenerates into the observed or empirical angular response and a noniterative solution is carried out.

REFERENCES

Anscombe, F. J. (1954). Discussion of Fisher (1954). *Biometrics, 10,* 141–144.

Bailey, N. T. J. (1951). Testing the solubility of maximum likelihood equations in the routine application of scoring methods. *Biometrics, 7,* 268–274.

Bartlett, M. S. (1954). Discussion of Fisher (1954). *Biometrics, 10,* 140–141.

Berkson, J. (1953). A statistically precise and relatively simple method of estimating the bioassay with quantal response, based on the logistic function. *J. Amer. Statist. Soc., 48,* 565–599.

Biggers, J. D. (1951). The calculation of the dose-response line in quantal assays with special reference to oestrogen assays by the Allen-Doisy technique. *J. Endocrinol., 8,* 169–178.

Biggers, J. D. & Claringbold, P. J. (1954). Optimum conditions for the local (intra-vaginal) action of oestrogens. *Aust. J. Biol., 7,* 118–139.

Biggers, J. D., Claringbold, P. J. & Emmens, C. W. (1954). A study of the variation in response of ovariectomized mice to the intravaginal and subcutaneous administration of oestrogens. *J. Endocrinol., 11,* 26–35.

Box, G. E. P. & Wilson, K. D. (1951). On the experimental attainment of optimum conditions. *J. R. Statist. Soc., 13B,* 1–45.

Campbell, R. C., Hancock, J. L. & Lord Rothschild (1953). Counting live and dead bull spermatoza. *J. Exper. Biol., 30,* 44–49.

Claringbold, P. J., Biggers, J. D. & Emmens, C. W. (1953). The angular transformation in quantal analysis. *Biometrics, 9,* 467–484.

Claringbold, P. J. (1953). The effect of metabolic inhibitors on the local action of oestrone and oestradiol-3:17β on the vagina of ovariectomized mice. *Aust. J. Biol. Sci., 6,* 657–664.

Cochran, W. G. (1938). Some difficulties in the statistical analysis of replicated experiments. *Emp. J. Exp. Agric., 6,* 157–178.

Finney, D. J. (1952a). *Probit Analysis.* 2nd ed. Cambridge University Press.

Finney, D. J. (1952b). *Statistical method in bioassay.* Griffin, London.

Fisher, R. A. (1922). On the dominance ratio. *Proc. Roy. Soc. Edin., 42,* 321–341.

Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Camb. phil. Soc.*, *22*, 700–725.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.*, *144A*, 285–307.

Fisher, R. A. (1946). A system of scoring linkage data with special reference to pied factors in mice. *Amer. Nat.*, *80*, 568–578.

Fisher, R. A. (1954). The analysis of variance with various binomial transformations. *Biometrics*, *10*, 130–139.

Fisher, R. A. & Yates, F. (1953). *Statistical Tables*. 4th ed., Oliver and Boyd, Edinburgh.

Potter, C. & Gillham, E. M. (1946). Effects of atmospheric environment, before and after treatment, on the toxicity to insects of contact poisons. *Ann. Appl. Biol.*, *33*, 142–159.

Rao, C. R. (1952). *Advanced statistical methods in biometric research*. Wiley, New York

LETTER TO THE EDITOR

To the Editor of *Biometrics*

Sir,

Dr. Hamaker's excellent paper (Vol. 11, p. 257) contains one implication that I hope will be considered false by most biometricians: I can write thus bluntly, because it is clear that Dr. Hamaker would be pleased to find himself mistaken on this point. He appears to suggest that in biology, and especially in agriculture—in contrast to industry—an analysis of variance is considered an adequate summary of the statistical examination of an experiment without any tabulation of means and standard errors.

An experienced statistician may often be able to judge from inspection of an analysis of variance what the main features of the interpretation of the experiment are, especially of course if he has himself computed that analysis. This will not blind him to the fact that well-arranged tables of means and their standard errors are the most important summaries of numerical information from any experiment. The analysis of variance, far from being a more sophisticated presentation of the information, is usually no more than a scaffolding needed in preliminary study of the data: it does not require inclusion in the ultimate report, since its duty is accomplished when it has indicated what tables of means need to be discussed and has provided estimates of standard errors. These functions are made clear in Hamaker's §8.

A contrary impression may sometimes be given by text-books and papers expository of statistical techniques. Naturally these give special attention to matters that may be unfamiliar to their readers, amongst which often are details of calculations of the analysis of variance for new experimental situations. Economy of space may prevent the authors from discussing how to prepare and present, for each set of data, good summary tables or diagrams, procedures with which they may assume—not always correctly—their readers to be familiar. Perhaps authors of text-books could usefully point out that while attempting to teach statistical techniques they cannot always be illustrating how best to present interpretative reports.

There are of course exceptions to my suggestion that the analysis of variance of an experiment does not need to be reported. Apart from its expository value, for example, it may give information on components of variance important to any discussion of the efficiency of the design and the possibility of improving it for subsequent work.

It would be unfortunate, however, if novices in biometric practice were to imagine that abandonment of tables of means, the natural and common-sense summaries of experimental data, in favour of tables of analyses of variance was an advance in statistical sophistication at which they should aim.

I do not question the appropriateness of Dr. Hamaker's tabulations for their purpose; I entirely agree that in other fields of application different styles of presentation may be needed, and that the style needs to be adapted both to the character of the data and to the statistical experience of the reader; I should deplore any tendency of statisticians to expect readers of their reports on experiments to comprehend 'sums of squares, degrees of freedom, and mean squares' *instead of* means of observations.

<div align="center">Yours truly,</div>

D. J. FINNEY

Department of Statistics
Marischal College,
University of Aberdeen,
Aberdeen, Scotland

25 October, 1955

QUERIES

QUERY: The following problem occurs frequently in some of **118** my studies and I would very much appreciate your opinion as to whether the following method of solution is correct.

Samples of unequal numbers of fish have been collected from 5 different locations in a lake. A one-way classification of analysis of variance indicates that their length is significantly different at the .01 level. Apparently it is the average size that is different since Bartlett's test on the variances is not significant. The problem is to establish whether any particular location or locations are reponsible for the difference while the others can be considered to belong to the same population. I believe we could find this out by calculating the fiducial limits of the means. The variance of the population would be the mean square within locations but I am uncertain what to divide this by for the calculation of the 95% semi-interval.

Should the mean square be divided by k_0 as per the formula on Page 234 of Snedecor's 4th edition?

This method would provide fiducial limits for all means in one calculation. Sometimes the numbers of fish in each sample vary greatly e.g., 35, 23, 7, 16, 41 = 122 fish for 5 locations. In this case would it be more correct to calculate a fiducial limit for each mean separately using the common mean square within population and their respective n's?

ANSWER: If a 95% fiducial interval is wanted, *separately* for the ith "location mean" μ_i this can, as you say, be computed from the usual formula as

$$\bar{x}_i \pm t_\nu(5\%) \cdot s \sqrt{k_i} \qquad (1)$$

where s^2 is the 'within location' mean square, ν its degrees of freedom and $t_\nu(5\%)$ the corresponding double tail % point of t, while k_i is the sample size for the ith location mean $\bar{x}_i$.

If a fiducial interval is wanted for each of the n location means μ_i then the above computation can of course be carried out for all locations. Computational labor may be saved by using an 'average' sample size k_0 in place of the separate k_i , but this appears hardly worth while. The question of *what* average of the k_i may be used arises. The answer depends on what properties it is desired the average fiducial interval should have. The formula for the average group size k_0 given by

Snedecor on p. 234 arises in the estimation of components of variance and it is difficult to conceive of a realistic reason why it should be used here. If it is desired that the average confidence coefficient should still be 0.95 this could be (approximately) achieved by using for the average sample size the value

$$\sqrt{k} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{k_i} \qquad (2)$$

An analogous problem arises when you are concerned with setting up a confidence interval for the *difference* between two particular location means $\mu_i - \mu_j$ which is given by the familiar formula

$$(\bar{x}_i - \bar{x}_j) + t_\nu(5\%) \cdot s\sqrt{\frac{1}{k_i} + \frac{1}{k_j}} \qquad (3)$$

Again, the problem of averaging may be posed here.

Finally the question arises as to how the fiducial intervals be used to decide the question as to which of the locations differ with regard to the fish-length. Presumably your idea was to call any two locations (i, j) with mean fish lengths μ_i , μ_j for which the confidence interval (3) does not include the value 0 different from one another. With such a procedure you would be making $n(n - 1)/2$ decisions about the $n(n - 1)/2$ differences and the question of your 'error-rate', i.e. your frequency of wrong statements, now becomes a little more complex than in the case of a single fiducial interval for a single pair of means.

A solution to your problem is given by Henry Scheffé in "A method for judging all contrasts in the analysis of variance," Biometrika 40:46–49 and 57–62 (1953). You will get a 5% overall experiment-wise error rate if you replace in (3) $t_\nu(5\%)$ by $\sqrt{(n-1) F}$. F is the 5% point of F for $n - 1$ numerator degrees of freedom and ν denominator degrees of freedom.

H. O. HARTLEY

CHANGE IN THE EDITORSHIP OF BIOMETRICS

With the December, 1955 issue of *Biometrics*, Miss Gertrude Cox will terminate her period of service as editor of *Biometrics*, a post which she has held since the first number came out, under the title *Biometrics Bulletin*, in February, 1945. On several occasions during recent years, Miss Cox has asked to be relieved of the editorship owing to the increasing pressure of her many other commitments in biometric activities. These requests were handled either by persuading Miss Cox to continue in the post, or in some cases, I fear, by pretending that they had not been heard. When the request was renewed in 1954, however, it was felt that the Society had imposed too long on Miss Cox's public-spiritedness, and that steps should be taken to seek a successor. A committee consisting of F. Yates (Chairman), D. Mainland and A. Linder was appointed to make a recommendation about a successor. After careful consideration of a number of possible nominees and sites, the committee unanimously recommended Dr. John W. Hopkins of the Division of Applied Biology, National Research Council, Ottawa, Canada, who kindly consented to take on the post after due approval by the Council. During the present year Dr. Hopkins has been serving as Associate Editor in order that a smooth and orderly transition of the work can be made. With the March 1956 issue he will assume full responsibility as editor.

I do not know whether members realize what a great debt they owe to Miss Cox. From an initial issue of 12 pages, she has built up *Biometrics* into a leading journal in its field. Her refereeing policy has been helpful and considerate to authors, while maintaining standards of high quality. She has devoted a great deal of thought and effort to obtaining expository papers and simple examples of the newer techniques that would be easily understood by biologists. Some of these efforts were unrewarded, for the number of competent people who are willing to prepare expository papers, or to deliver them after having agreed, is distressingly small. Many of the special issues of *Biometrics* that have been in wide demand were planned and stimulated by Miss Cox. She has kept the publication on a sound financial basis through the troublesome early years, despite some disappointments in securing financial aid that had been anticipated. The steadily growing requests for complete sets of the journal from libraries, institutions and individuals, is a testimony to the high regard in which it is held. I want to

take this opportunity on behalf of the members of extending to her our warmest thanks.

Thanks are also due to Mrs. Sarah Porter Carroll of the Institute of Statistics, North Carolina State College, who has served diligently and competently as managing editor of *Biometrics*, and has relieved Miss Cox of much of the time-consuming labor that is unavoidable in editing a journal. North Carolina State College has contributed generously by making available secretarial help, space and facilities.

Dr. Hopkins, who is Chairman of the Finance Committee and has served on the Council, will be well-known to most members. I am sure that he can count on the full cooperation of members in maintaining the high quality that has been set.

W. G. Cochran
President

325 PAUL E. FIELDS (School of Fisheries, University of Washington). **Factorial Designs and the Guidance of Downtsream Migrant Salmon and Steelhead Trout.**

This study is a part of the Columbia River Fisheries Engineering Investigation and Research Program sponsored by the North Pacific Division, Corps of Engineers. Probably because of difficulties encountered in rearing anadromous salmon, there was but little basic information about their sensory abilities and their behavior patterns when it suddenly became necessary to find some effective guiding stimulus. The avoiding response to light seemed to offer the most promise of success.

In the first series of factorial experiments, the reactions of a total of 90 different groups of 25 one year old silver salmon, each given four trials, was obtained to a light barrier with three angles and three light intensities, in water of four velocities and three depths. In general, the number of fish entering the lighted area was significantly reduced as the angle of the barrier was made smaller, the intensity of the light was increased, and the velocity was decreased. The F for depth was not significant. In a second experiment, the responses of a total of 72 groups of steelhead trout, chinook and silver salmon were compared on two barrier angles, two light intensities, and two water velocities. The findings of the previous experiment were confirmed and the range extended. In addition, a species difference was established with steelhead trout being the most sensitive to light. In a third experiment, the reactions of a total of 48 groups of 50 each of steelhead trout, chinook and silver salmon were compared with respect to chain barriers with two angles, two different spacings and two water velocities. The only significant F was between species.

As the success in guiding has increased, the normality of the data has decreased, making the application of the usual parametric methods more questionable.

326 D. G. CHAPMAN AND R. PYKE. (University of Washington). **The Statistical Theory of Some Migration Population Models.***

The problem considered is that of estimating the parameters associated with the migration of individuals between two areas (A_1 and A_2). The populations studied are comprised of two classes, the X-type and

*Work done partly under the sponsorship of the U. S. Office of Naval Research.

the Y-type individuals. Let $X_i(Y_i)$ be the number of X-type (Y-type) individuals in A_i before migration. We assume that these parameters are known and that $P_1 \neq P_2$ where $P_i = X_i/N_i$ and $N_i = X_i + Y_i$, $(i = 1, 2)$. Define M_{xi} as the number of X-type individuals migrating from A_i to A_{3-i}, and define M_{yi} similarly. Set $M_i = M_{xi} + M_{yi}$. The following models have been studied.

Model I: Assume

(a) migration occurs in one direction only, from A_1 to A_2, say;
(b) M_{x1} is a random variable, distributed as $b(M_{x1} : M_1 , P_1)$
(c) a sample of size n_2 is taken in A_2 after migration in which X_2 X-type individuals are observed.

The estimator of M_1,

$$
\hat{M}_1(x_2) = \begin{cases} N_1 & \text{for } x_2 \text{ between } j \\ & \text{and } n_2(X_1 + X_2)(N_1 + N_2)^{-1} \\ \\ N_2 \dfrac{n_2 P_2 - x_2}{x_2 - n_2 P_1} & \text{for } x_2 \text{ between } n_2 P_2 \\ & \text{and } n_2(X_1 + X_2)(N_1 + N_2)^{-1} \\ \\ 0 & \text{otherwise} \end{cases}
$$

where j is 0 or according as $P_1 - P_2$ is negative or positive, is derived, studied and the conditions for its reliability outlined. Approximate formulae for the expectation and variance of it are given. It is shown that for large parameter values, x_2 is approximately distributed as

$$
b\left(x_2 : n_2 , \frac{X_2 + MP_1}{N_2 + M} \right)
$$

Model II: Assume (a) and

(d) simultaneous samples of size n_i are taken in A_i after migration in which x_i X-type individuals are observed $(i = 1, 2)$.

The M.L. estimators of M_1 and M_{x1}, for $z_i = x_i/n_i$, are

$$
\tilde{M}_1(z_1 , z_2) = \frac{N_2(P_2 - z_2) + N_1(P_1 - z_1)}{z_2 - z_1}
$$

$$
\tilde{M}_{x1}(z_1 , z_2) = \frac{z_1 N_2(P_2 - z_2) + N_1(P_1 - z_1)}{z_2 - z_1}
$$

of which the asymptotic properties are studied.

Model III: Assume (a) and

(e) the migrants are distinguishable from the residents,

(f) a sample of size n is taken in A_2 after migration in which r_i X-type and s_i Y-type individuals are observed, the subscript denoting the area in which the individuals were before migration

The M.L. estimators of M and M are

$$\overline{M}_1(r_i, s_i) \doteq \frac{(r_1 + s_1)N_2}{r_2 + s_2}, \qquad \overline{M}_{x1}(r_i, s_i) = \frac{r_1 N_2}{r_2 + s_2},$$

of which the asymptotic properties are studied.

Model IV: Assume (d) and

(g) migration occurs in both directions.

In this case, the same estimators are obtained as in Model II, except that now negative values of these functions make sense.

327 JOSHUA L. BAILY, Jr., Sc.D. (San Diego, California). **Variation of the** *Pecten Gibbus* **Complex.**

This is a repetition of Davenport's "Quantitative Studies on the Evolution of *Pecten*" made about half a century ago, to see what changes, if any, have taken place in the meantime.

Pecten is a bivalve mollusc, and like other bivalve molluscs has two valves which are organically right and left. But the species in this investigation when at rest, and also when swimming have changed their ancestral orientation, the primitively right and left valves having become in both cases the lower and upper valves respectively, and the functional right and left halves being organically anterior and posterior. The question then arises as to whether in general functional considerations are more influential than organic relationships in determining correlations.

Davenport also concluded that species from the Pacific coast are in general more variable than closely related species from the Atlantic coast. This differential variability he concluded could be attributed to the differences in the geological history of the two coasts. But Davenport's conclusions were based upon measurements of dimensions, and the difference in variability in size might be more simply explained by assuming a greater number of age groups represented in the Pacific series of specimens. Coefficients of variation of ratios might conceivably have a different result, since they would indicate variability of shape rather than of size.

R. F. TATE AND R. L. GOEN. (University of Washington).
328 **Minimum Variance Unbiased Estimation for a Truncated Poisson Parameter.**

A problem of current interest, especially in public health work, is that of estimating the parameter of a Poisson distribution which is singly truncated on the left. An MVUE is found for the general case, with simple expressions resulting for the two important special cases of truncation away from zero, zero and one. Results proceed from considerations involving sufficient statistics.

329 HERBERT D. KIMMEL. (University of Southern California). **The Reliability of Categorical Qualitative Judgments.**

While the problem of estimating the reliability of quantitative data such as test scores or qualitative data which may be artificially quantified along one qualitative dimension has been met adequately by application of one of several methods based on score-variations, no generally applicable method has been devised to estimate the reliability of a qualitative rating or sorting schema which cannot be artificially quantified.

This paper proposes a new method for estimating reliability in such situations which is based on the proportion of agreement obtained among several judges and which takes into account the proportion of agreement which would be expected to obtain by chance alone. The reliability estimate obtained is logically analogous to internal consistency type measures, on the assumption that each individual judge acts as a separate item in a test. The method gives values ranging between zero and unity.

In addition to providing a reliability estimate for the whole schema, the method may be used to obtain separate reliability estimates for the separate qualitative categories. It should be noted, however, that these separate category-estimates are somewhat dependent upon the number of times the category was used. Also, the method is not recommended in situations which may be artificially quantified with reasonable justification.

DAVID A. GRANT. (University of Wisconsin). **Statistical**
330 **Tests in the Comparison of Curves (by Means of Orthogonal Components of Trend).**

This paper extends the well-known Alexander Trend Analysis procedure in two respects. The Alexander procedure applies where

there are a series of scores, obtained by repeated trials on the same Ss, in two or more groups. It provides for comparison of the groups in terms of: (a) their mean differences; (b) differences in the linear components of the group trends; and (c) differences in the pooled higher-order components of the group trends. The present procedure is more analytic in that: first, the groups may be compared separately in terms quadratic, cubic, and any further orthogonal components of the trends; and secondly, if the groups form an orthogonal array, *e.g.*, rows and columns, the row, column and interaction variation may be examined separately for linear, quadratic, cubic, etc. differences.

The tests are obtained by constructing covariance terms by means of the orthogonal polynomials. Using Cochran's theorem, it is easily shown that, with a mathematical model, linear in the orthogonal polynomial components, with normal, random, and equal error variation, the separate component tests conform to the F distribution. A routine method of calculation of the sums of squares has been worked out with suitable checking procedures.

The procedure is limited to cases where the intervals between trials or levels of the corresponding independent variable are equal on a linear, logarithmic or similar scale. It has proved most valuable in our laboratory for comparing experimental curves separately with respect to slope, curvature, inflections, and the like. It is not particularly efficient when the curves are expected to follow exponential or other transcendental functions.

331 PHILIP R. MERRIFIELD. (University of Southern California). **Quantification of Ordering Behavior.**

Ordering behavior occurs in several contexts. It is defined here as the process of arranging objects or situations, or verbal definitions thereof, in the order most appropriate with reference to a criterion. The criterion may or may not be stated explicitly, but in general it has the characteristics of (a) a time continuum, (b) a spatial arrangement, or (c) an hierarchical system. These three are contextual cases of what might be labelled as "logical" arrays.

Two major alternative hypotheses as to the nature of the ordering process are entertained. Under the first, it is suggested that the examinee treats the set of stimuli as a whole, transforming the entirety into a new array by a process that is primarily "unitary;" subhypotheses deal with minimizing the "error space" in what correspond to the two-dimensional and three-dimensional cases. Under both subhypotheses, double and triple interaction effects are considered.

Under the second major hypothesis, it is suggested that the examinee deals separately with the individual elements in the problem, or at most considers the individual elements in pairs.

On the premise that scoring systems based on these hypotheses will disclose differences in total scores sufficient to support a decision in favor of one or the other, seven scoring systems are derived.

Selected results from a factor-analytic investigation of planning abilities, carried out by R. M. Berger, J. P. Guilford, and P. R. Christiansen, and from a smaller separate study carried out by the writer, are discussed with reference to the hypotheses concerning the nature of the ordering process.

332 J. A. GENGERELLI. (University of California, Los Angeles). **A Method of Constellation Analysis.**

The intent of the method is to determine whether an assemblage of objects is comprised of mutually exclusive sub-classes or whether it constitutes a single continuum. The problem arises in a variety of contexts, viz.: the question of (a) physical types, (b) personality types, (c) psychiatric nosologies, (d) psychological factors, (e) cultures.

The central concepts are those of "neighbors," "neighborhood," "distance," and "constellation." A constellation is defined and said to exist if a set of objects are mutually neighborly; two or more objects are mutually neighborly when all are at no greater distance one from another than a certain critical value s. The critical value s for any given universe of discourse is defined as the average of the distances separating every pair of objects constituting that universe. In certain contexts, however, a constellation is not determined by distance but by the fact of concomitance. Thus, a set of objects is said to be a constellation if the objects are concomitant, i.e., all occur if any one occurs.

A method for isolating constellations is described and applied to representative problems.

333 HARRY H. HARMAN. (The Rand Corporation). **Some Observations on Factor Analysis.**

This paper is non-mathematical and expository in character. A brief review is given of the origin and growth of factor analysis, delineating (1) its use in the formulation of psychological theories of human abilities and traits, and (2) its consideration as a branch of general statistical technique. The question is raised regarding the basis of choice of a particular factor solution out of the infinite possi-

bilities that arise in reducing a given matrix of correlations. Preferred types of solutions are enumerated, with an indication of the extent to which the two general principles—statistical simplicity and psychological significance—affect each type. Some of the problems that have plagued factor analysts during the past half century are pointed up. Finally, an appraisal of the present status of the subject and a prognosis for the future is made.

334 J. W. FRICK. (University of Southern California). **The Effect of Varied Interpolated Stimuli upon the Time Order Error.**

It was hypothesized that varied stimuli, interpolated between a constant stimulus and certain variable stimuli, would have a differential effect upon the judgment of the variable stimuli as compared with the constant. It was further hypothesized that, as the interpolated stimuli approached the constant in size, reinforcement of the latter would occur, resulting in a lesser number of negative TOE's.

Twenty S's made a total of 1600 judgments of 4 randomly-presented variable stimuli in comparison with a constant stimulus, under four conditions of interpolated stimuli. All stimuli were black lines projected tachistoscopically on a constantly-lighted white screen, and varied in projected length from 18.8 to 21.2 inches. The constant was 20 inches in length.

As expected, the number of negative TOE's exceeded the number of positive errors, but were reduced to a frequency less than that to be expected by chance. An analysis of variance disclosed no significant differences between judgments made under the varying stimulus conditions. This may indicate that all interpolations had a reinforcing effect upon the constant stimulus, since the largest interpolated stimuli varied only plus-or-minus 2 j.n.d.'s (1.2 inches) from the constant. (The author is indebted to Dr. Harvey F. Dingman for the analysis of the data.)

335 DR. LLOYD A. LIDER. (University of California at Davis). **A Group of Long-term, Perennial and Non-replicated Root-stock Trials.**

Data on production, fruit quality and vigor have been taken on a group of cooperative grape rootstock trials established in the grape growing areas of the coastal counties of California. These trials, using experimental phylloxera resistant rootstocks, were set up over the

last twenty years, and annual measurements have since been taken. From seven to ten rootstocks appear in each non-replicated trial. They have been planted under a wide range of environmental conditions and have represented most of the commercial scion varieties of the area. From the data on hand it is possible to draw certain conclusions concerning the influence of the rootstocks on the behavior of the scion varieties. An interpretation of these data are presented as well as general conclusions concerning the usefulness of and difficulties presented by a study of this nature.

CLYDE STORMONT. (University of California at Davis).
336 **Estimates of Frequencies of B Alleles in Three Breeds of Dairy Cattle.**

Computational and theoretical difficulties in estimating the frequencies of even such well-adapted genes as those controlling blood groups trace to the frequent operation of extensive allelic series and the necessity of taking family relationship into account. Our model here is the 100 or more alleles which control the exceedingly complex B system of bovine blood groups—a system which has maximum utility in present and projected application along such diverse lines as medicolegal problems, evolution theory, clinical genetics, animal breeding plans, immunology, genetic theory and so on. It has been known for a number of years that many of these B alleles are in principle almost breed specific but precise statements as to their frequencies within breeds have not been made. Marked differences in frequencies between lines within breeds pose a problem of obtaining samples that are representative of the various breeds. The computational problem could be reduced to its simplest form by developing serological reagents capable of differentiating all of the 5,050 or more phenotypes in the B system. Considerable progress towards this goal has been made and consequently, efficient estimates with small errors due to misclassification of some of the phenotypes can be obtained by simply computing the ratios of the actual allele counts. Assuming there are no sex differences in regard to frequencies of B alleles, the sampling of bulls in various semen producing businesses throughout the United States would seem to provide one of the best methods of obtaining samples that are representative of the various dairy breeds. "Preliminary" estimates of the frequencies of B alleles in Guernsey, Jersey and Holstein-Friesian breeds in U. S. are based on samples of 200 Holstein-Friesian, 80 Jersey and 80 Guernsey bulls.

337 HERMAN RUBIN. (Stanford University and University of California at Berkeley). **Axiomatization of Genetics.**

Given a theory of genetics (such as Mendelian inheritance without mutations), it is desired to construct a mathematical system which has the properties the geneticist ascribes to his theory. One wishes to list a number of mathematical entities and a number of axioms connecing them: also to call the mathematical model an axiomatization of genetics, it is necessary to correlate genetic terms with those mathematical entities, called *primitive notions*, which are not explicitly defined. In addition, genetic terms might correspond to defined entities; for example, if the relation *immediate ancestor* is a primitive notion, the relation *ancestor* could be a defined notion. In most genetic models, especially those on a cellular basis, a large number of primitive notions is required. For example, in Mendelian inheritance, *haploid cell, diploid cell, genetic map, genetic description*, and the relations of *mitosis, meiosis*, and *fertilization* would normally be some of the primitive notions. Typical examples of axioms are *A genetic map is a subset of Euclidean two dimensional space with the x-coordinate an integer*, and *If A is related to B by mitosis, the genetic descriptions of A and B are identical*. An example of a statement which would not belong to a genetic axiom is *During mitosis, the new chromosomes are drawn to their respective nuclei*.

Australasian Symposium on Biometry, University of Melbourne Australia, Monday, 22nd August, 2:15 p.m.

338 R. T. LESLIE. (University of Melbourne). **A Statistical Approach to the Physiological Problem of Thresholds.**

For a stimulus of given intensity I, a certain change ΔI is necessary for the difference to be detectable; the classical Weber-Fechner law is to the effect that within a certain range $\Delta I/1 = $ constant.

It is known that a stimulus is translated into electrical pulses in the receptor nerve, the frequency of the pulses being related to the stimulus intensity. To distinguish two stimuli as of different intensity the discriminating organ must compare two pulse rates, known to be each subject to random fluctuations. Assuming storage of information, the problem is analagous to the comparison of a pair of samples, where sampling is over (n) units of time, and discrimination should improve

in proportion to $1/\sqrt{n}$. A similar argument may be applied to sampling of stimulated area (visual and tactile stimuli), discrimination again improving with increase in area.

To explain why $\Delta I \neq 0$ for sufficiently protracted or intensive stimuli ($n \rightarrow \infty$) random noise in the central nervous system may be assumed, the noise adding a constant and irreducible amount to the variance of the difference between the impulse rates.

Some further predictions from this theory have been verified experimentally.

339 E. J. WILLIAMS. (Division of Mathematical Statistics, C.S.I.R.O., Melbourne). **Sidelights of Sampling Surveys.**

Some of the problems arising in the conduct of sampling surveys and the interpretation of their results are discussed. A survey seldom works out in the field as originally planned.

Some experiences with the New Guinea Census of Native Agriculture are described. How should one treat a sample village which is found to have—
   (i) disappeared,
   (ii) migrated to another district, or
   (iii) split up into several villages?

The effects of methods of sampling and of ascertainment of attributes on the method of estimation is considered. Cases arise where the probability of inclusion of an item is—
   (i) proportional to size,
   (ii) inversely proportional to size,
   (iii) modified by the method of sampling adopted.
Each case requires a different form of estimate.

The problem of bias in estimates from small samples is important for sampling surveys. Methods of reducing bias are discussed.

340 G. S. WATSON. (Australian National University, Canberra). **Missing and Mixed-Up Values in Contingency Tables.**

In the analysis of variance, the problems of missing and mixed-up plots have well-known solutions. However the same problems may arise in the analysis of contingency tables. It is shown in this paper that they may be solved by an application of the method of maximum likelihood.

341 P. J. CLARINGBOLD. (University of Sydney). **Discriminant Analysis in the Interpretation of Semi-quantal Data.**

Intermediate between quantal responses and the fully quantitative or graded responses are responses which may be described by the term semi-quantal. These may simply be regarded as a generalization of the quantal response to the case where more than two response classes are recognised. In the past, data of this type from insecticidal studies were always reduced to a quantal form by grouping of response classes. This was carried out because of mathematical difficulties associated with the extension of probit analysis to semi-quantal response. In his "Statistical Methods" Fisher shows how discriminant analysis may be employed to derive scores for various types of response. The method described is extended in this paper and illustrated by an example taken from studies on the action of oestrogens.

342 (Mrs.) G. L. RICHARDSON and F. E. BINET. (University of Melbourne). **Discriminant Analysis on Species of the Genus** *Murravia* **(Brachiopoda, Tertiary and Recent).**

Specimens of both *M. triangularis* and *M. lenticularis*, obtained from different localities (all restricted to the "lower Aldingan" of Tate), exhibit little or no variation in specific characters. Statistical analysis on samples of *M. triangularis* shows that there is no evidence against over-all homogeneity in these samples.

M. catinuliformis shows a considerable range of variation both within and between different collections which are derived from deposits over a wide stratigraphic range (Janjukian to Cheltenhamian Stages). These collections are empirically divided into three groups which, it is suggested, may form a basis for future specific or sub-specific distinction.

A statistical analysis is based on the logarithms of (1) breadth, (2) distance from the posterior point of the pedicle valve to the line of greatest breadth, (3) distance from the anterior tip of the pedicle valve to the line of greatest breadth.

A size factor is first defined as the linear combination with co-efficients equal to the reciprocals of the estimated standard deviations. General linear combinations are then formed, and from those which appear to be uncorrelated with the size factor, two are chosen, one maximizing and the other minimizing the variation between collections relative to that between individuals. The variation is significant in both directions. These optimal combinations are significantly better discriminators than those derived from indices suggested by the morphology of the Genus.

NOTE

During the past few years I have been repeating the investigation of variability of *Pecten* made by Davenport half a century ago.

One of the valuable features of this work was the extensive bibliographies, with references to parallel work in other fields. The interest and value of the report of my work would be greatly enhanced if I could add to it references to similar investigations carried on since Davenport's work was published.

Davenport found the Pectens of the Pacific coast to be more variable than those of the Atlantic. I agree with this belief, but I would like to know if it holds true for the other forms of life, either animal or vegetable. Are Pacific coast species always more variable than geminate species from the Atlantic slope?

He published a list of investigations of the correlation between the right and left paired structures of bilaterally symmetrical organisms. I would like references to work since 1905 in which such correlations are reported.

<div align="right">

Joshua L. Baily, Jr.
4435 Ampudia Street
San Diego 3, California

</div>

ANNOUNCEMENT

Research proposals directed to the Division of Biological and Medical Sciences of the National Science Foundation of the U.S.A. will be received at any time. The proposals on research projects to begin in June or September 1956, will be reviewed during March. These proposals should be received by the Foundation prior to February 1, 1956.

Projects in the areas of anthropology, human ecology, functional archaeology, experimental social psychology, and demography are included in the Division's program.

NEWS OF MEMBERS

Dr. R. Lowell Wine, who received his B.A. degree at Bridgewater College, major in Mathematics, and his M.A. degree at the University of Virginia, major in Mathematics, and his Ph.D. in Statistics at the Virginia Polytechnic Institute, has joined the staff of the Department of Statistics of the Virginia Polytechnic Institute as Associate Professor. He has previously taught at the University of Virginia, Amherst College, University of Oklahoma, and Washington and Lee University. He was

a student in Statistics at the University of Michigan for two years prior to coming to Virginia Polytechnic Institute. Dr. Wine will do both teaching and research.

Dr. Rudolf J. Freund has joined the staff of the Department of Statistics as Associate Professor at the Virginia Polytechnic Institute. Dr. Freund did his undergraduate work at North Carolina State College, received his Master's degree in Statistics and Economics at the University of Chicago, and his Ph.D. degree in Statistics at North Carolina State College. Dr. Freund will do both teaching and research.

Helen Bozivich, research associate in the Statistical Laboratory of Iowa State College, has accepted a position at Purdue University as assistant professor in its Statistical Laboratory beginning July 1, 1955. She was awarded the degree of Doctor of Philosophy in statistics by Iowa State College in June, based partly on a dissertation, "Power of test procedures for certain incompletely specified random and mixed models."

John F. Hofmann, chief statistician of the Naval Ordnance Laboratory at Corona, California, was granted a Ph.D. degree in statistics by Iowa State College June 1955. His dissertation concerned "Life testing in controlled environmental conditions."

Bernard Ostle has been appointed Professor of Mathematics and Statistics at Montana State College.

In December 1954, the University of Melbourne created a full chair in Statistics and appointed Associate Professor **M. H. Belz** as the first Professor. This is the first full chair in Statistics at a teaching University in Australia. **Dr. G. S. Watson** of the University of Melbourne has accepted an appointment as Senior Research Fellow in the Department of Statistics at the Australian University in Canberra, effective in March of this year. The new Senior Lecturer at Melbourne University will be **Dr. H. A. David** of the Department of Mathematical Statistics, C.S.I.R.O.

John W. A. Brant, formerly Agricultural Officer of the Food and Agriculture Organization of the United Nations (1953–1955), now Specialist of the Universidad de Guayaquil y Universidad de Idaho en Programa Cooperativo para el Progreso de las Ciencias Agropecuarias, has been honored November Eighteenth by nomination to Professor, Facultad de Agronomia y Veterinaria during a Sesion Solemne de la Facultad de Agronomia y Veterinaria, Universidad e Guayaquil, Guayaquil, Ecuador—Octogésimo octavo aniversario de su Fundación. He has launched a research program in poultry nutrition, which is to be continued concurrently with research programs in animal physiology and genetics.

INTERNATIONAL BIOMETRIC SYMPOSIUM ON "THE ROLE OF BIOMETRIC TECHNIQUES IN BIOLOGICAL RESEARCH"

The second International Biometric Symposium, on the role of Biometric Techniques in Biological Research, met under the sponsorship of the Biometric Society, a section of the International Union of Biological Sciences, and under the auspices of the University of São Paulo represented by its Seminario de Estatistica. The Symposium was convened by the President of the Society, Professor W. G. Cochran, shortly after 10 a.m. on July 4. He introduced the Secretary of Agriculture for the State of São Paulo, Dr. R. Cruz Martins, who welcomed the Symposium to Brazil and Campinas and wished it success in its deliberations. In reply, President Cochran thanked Dr. Cruz Martins for his country's hospitality and good wishes. He then delivered his presidential address on the 1954 poliomyelitis trial in the United States, as an illustration of the critical role played by biometry in solving a major public health problem.

Following his address, President Cochran called upon Secretary Bliss for a summary of recent Society activities. The Secretary noted that proceedings of the Third International Biometric Conference of the Society, at Bellagio, Italy, in September, 1953, had appeared the following December in Biometrics, which has also published a number of the papers presented at the Conference. Under the editorship of Professor G. M. Cox, Biometrics had attained world-wide repute, but after 11 years, she had asked to be relieved of this post. The Society was fortunate in having obtained Dr. John W. Hopkins of Ottawa, Canada, as her successor, beginning with Volume 12.

In sponsoring the present Symposium, the Society was also acting in its capacity of Biometric Section of the International Union of Biological Sciences. The Union had allotted $2000 toward the travelling expenses of principal participants and $500 towards the publication of our proceedings. All papers read during the Symposium may appear in abstract in Biometrics and a number of them in full. At the XII General Assembly of the IUBS in Rome this past April, the section was represented by Drs. L. L. Cavalli-Sforza and A. Vessereau. Again with financial assistance from the IUBS, the Society (or Section) was

sponsoring an international Biometric Seminar to be held in Varenna on Lake Como, Italy, on September 7 to 23 of this year.

He noted further that many biometric meetings have been sponsored by national and Regional subdivisions of the Society in all parts of the world, so that in 1954 alone more than 20 of these had been reported. The Society now has a membership of 1300 with eight organized Regions and Regions authorized in Japan and also in Brazil when a sufficient number of members has been enrolled. The Brazilian members hoped to reach this goal before the end of the present Symposium. An organization meeting had been scheduled on the program. Following announcements by Dr. C. G. Fraga, the morning session was adjourned.

The first scientific session on Monday afternoon concerned bio-metrical genetics with the papers listed in the Scientific Program. A group of geneticists at the afternoon program met again in the evening to hear an address by H. Kalmus.

Experimental design was the subject of both sessions on Tuesday, July 5, the papers in the morning dealing especially with perennial crops. At 5 p.m. the participants in the Symposium were received in a pleasant reception by the Mayor of Campinas. In the evening, officials of the Biometric Society met with Brazilian scientists to discuss the formation of a Brazilian Region. At the end of a lively discussion, a committee was named to prepare a tentative set of Regional by-laws.

The following morning the Symposium moved by bus to Piracicaba for the day, a distance of 60 kilometers, where it continued with a panel discussion on experimental designs for perennial crops at the Escola Superior de Agricultura "Luiz de Queiroz", the Agricultural College of the University of São Paulo. Following a delightful luncheon of typical Brazilian dishes as guests of Professor Brieger's Department of Genetics, members of the Symposium visited the School and its associated Experiment Station and then were taken on a tour to a nearby sugar mill, to a paper factory, which used the bagasse from the sugar cane as raw material, to the Sugar Experiment Station, where its research program was summarized briefly, and finally to a social club in Piracicaba for coffee, refreshments and music before returning to Campinas by bus.

The morning session on Thursday, July 7, concerned the statistics of animal feeding experiments. The committee on by-laws for the Brazilian Region met at lunch and continued their work through the afternoon. Sampling techniques was the subject of the afternoon session of the Symposium. Members of the Council and other officers of the Society attending the Symposium met at dinner for informal discussions. In the evening, the Tenis Clube of Campinas entertained

the members of the Symposium at a social gathering, highlighted by a superb exhibition of native dances by members of the Club.

On Friday, July 8, the morning was devoted to a visit to the Instituto Agronômico which has its headquarters in Campinas. Founded in 1881 and now supported entirely by funds from the State of São Paulo, it is the oldest agricultural research institution in Latin America. In addition to its main Experiment Station of 2000 acres, located near Campinas, the Institute has 19 branch Stations in various ecological areas of the State and 31 technical sections grouped into four divisions of agronomy, biology, soils and technology, and experiment stations. The afternoon session presented three papers and a panel discussion on bioassay. Following the panel discussion, the Brazilian members of the Society completed the formation of the Brazilian Region. In the evening the members of the Symposium were entertained at a barbecue and dance by the Sociedade Hipica at its ranch near Campinas.

The last scientific session of the Symposium, on the morning of July 9, concerned medical statistics. At the final meeting the following resolutions were adopted unanimously.

"The officers and members of the Biometric Society meeting at Campinas, 4–9 July 1955 extend their sincere appreciation to the Committee on Arrangements: Chairman, F. G. Brieger, and his committee: P. Mello Freire, F. Pimentel Gomes, A. Groszmann, A. M. Penha, W. L. Stevens and C. G. Fraga Junior, Executive Secretary. Our special regards are due to C. G. Fraga Junior for his patience and kind consideration of our individual and collective problems.

It is resolved that our thanks be expressed to Director C. A. Krug and his staff at the Instituto Agronômico; to Director E. R. Nobre, Escola Superior de Agricultura "Luiz de Queiroz"; and to Geneticist F. G. Brieger and his staff. The field trips provided an opportunity for scientific activities combined with relaxation.

It is further resolved that we express our indebtedness to the Brasilian scientists for providing an extensive sampling of national foods and drinks with their ever present hospitality.

We express thanks to the clerical staff for efficient and untiring efforts in our behalf, and to the photographer for his records of serious and festive events."

The Secretary reported that the Society would hold its next international congress in Canada in the summer of 1958, to consist of the Fourth International Biometric Conference and a Symposium with IUBS support on some phase of biometrical methods in genetics. Their time and place would be synchronized with the Xth International Genetic Congress, which was also scheduled for Canada in the same

summer.   Following announcements by Dr. Fraga, the Symposium was declared adjourned by President Cochran.   In the afternoon, those remaining after the Symposium were taken for a tour of a dairy and coffee farm near Campinas, and on the following day to Santos, the port for São Paulo on the Atlantic, both delightful excursions.

SCIENTIFIC PROGRAM

(The program of the Symposium was arranged by an Organizing Committee consisting of W. G. Cochran (Chairman), C. I. Bliss, F. G. Brieger, F. J. Crow, D. J. Finney, C. G. Fraga, J. A. Rigney and P. V. Sukhatme.

July 4. 10 a.m. Presidential address by W. G. Cochran—The 1954 trial of the poliomyelitis vaccine in the United States. (see page 527).

3 p.m. Biometrical Genetics. Chairman: F. G. Brieger. Sir Ronald Fisher—The contribution of biometry to plant breeding. E. R. Dempster—Genetic models in relation to animal breeding. F. G. Brieger—Behavior of autogamic populations and heterotic genes. H. Kalmus—Some genetical consequences of cyclomorphosis.

July 5. 9 a.m. Experimental Designs for Perennial Crops. Chairman: W. G. Cochran. S. C. Pearce—The specific problems of experimental design and technique in perennial crops. E. Amaral—The estimation of missing plots in perennial crops. A. Conagin and C. G. Fraga—Design and analysis of coffee experiments. F. Pimentel Gomes —Methods of describing crop response to fertilizers in perennial crops.

2 p.m. Experimental Designs. Chairman: G. Darmois. G. M. Cox—Recent advances in experimental designs with particular reference to estimating responses to rates of application. W. J. Youden—Design of experiments in the physical sciences.

July 6. 10 a.m. Panel Discussion on Experimental Designs for Perennial Crops. Chairman: W. L. Stevens.

July 7. 9 a.m. Statistics Applied to Animal Feeding Experiments. Chairman: B. B. Day. P. G. Homeyer—Technique and sources of variation in animal feeding experiments. G. L. da Rocha—Grazing experiments in the state of São Paulo. G. O. Mott—The grazing trial for measuring the output of pastures. A. Linder—On a particular kind of grazing experiment.

1:45 p.m. Sampling Techniques. Chairman: A. M. Flores. M. H. Hansen and J. Steinberg—Control of errors in surveys. P. V. Sukhatme—Sampling techniques for estimating the catch of sea fish. J. Nieto de Pascual—National morbidity survey in Mexico. W. L. Stevens and S. Schattan—The sampling of coffee for forecasting harvests. E. Cansado—Sampling without replacement from finite populations.

July 8. 1:30 p.m. Bioassay. Chairman: W. J. Youden. C. I. Bliss—Confidence limits for measuring the precision of bioassays. D. J. Finney—Cross-over and single-subject design for 4-point assays. O. G. Bier and P. Mello Freire—Application of bioassay to complement fixation. By title, M. Masuyama and M. Hatamura—Recent advances in biometry in Japan.

4:30 p.m. Panel Discussion on Bioassay. Chairman: A. Linder.

July 9. 9 a.m. Medical Statistics. Chairman: G. Rasch. J. O. Irwin—The study of the physiological effects of hot climates. J. Manceau—Application of the covariance analysis to the comparative study of two anthelmintics. A. E. Brandt and G. H. Fletcher—Design of a clinical investigation of very high voltage sources in the radio-therapy of cancer. A. Vessereau—Utilisation de l'analyse discriminatoire pour un diagnostic medical.

REGISTRATION

The 98 participants in the Symposium represented 17 different countries and of this group 57 are members of The Biometric Society. The following list gives the participants by countries: *Brazil (State of São Paulo)*—A. Sousa Quieroz do Amaral, Fernando Andreasi, H. Antunes Filho, H. Vaz de Arruda, S. Correa de Arruda, Elza S. Berquo, Otto Bier, A. A. Bitancourt, F. G. Brieger, L. de Freitas Bueno, Claudio Carvalheira, Armando Conagin, Candida H. T. M. Conagin, D. Mattas Dedecca, M. S. Dias, A. D. Netto, Constantino G. Fraga, Edison Galvao, J. F. Harrington, S. B. Henriques, Warwick Estevam Kerr, C. A. Krug, R. Aguiar da Silva Leme, Walter Leser, F. F. Manzoli, R. Franco de Mello, P. Mello Freire, A. J. Teixeira Mendes, Jose Mitidieri, Antonio Morales, A. Martins Penha, F. Pimentel Gomes, A. Mendes Peixoto, Humberto Rangel, G. L. Rocha, Victoria Rossetti, Anesiades Salati, W. R. A. Schottler, M. Rocha e Silva, J. Soubihe Sobrinho, W. L. Stevens, Odette C. Toledo, G. Pinto Viegas, Mario Zaroni; *Brazil (Other States)*—Edilberto Amaral, G. Garcia Duarte, Americo Groszmann, Jose Grossman, Virgilio Libonati, Ruben Markus, J. M. Pompeu Memoria, R. Meirelles de Miranda, J. Soares Neves, A. Figueiredo Penteado, F. Costa Pereira, J. B. de Barros Pimentel, A. R. da Silva, G. A. Drummond, J. N. Manceau, A. Garcia de Miranda Neto, Erik Smith, Estavam Strauss. *Argentina*—M. Guibourdenche de Cabezas. *Bolivia*—J. H. Jimenez. *Chili*—Enrique Cansado. *Colombia*—B. Romero Rojas. *Costa Rica*—Mario Gutierrez. *Denmark* —G. Rasch. *El Salvador*—Floyd R. Olive. *France*—G. Darmois, A. Vessereau, P. E. Vincent. *Great Britain*—D. J. Finney, R. A. Fisher, J. O. Irwin, H. Kalmus, S. C. Pearce. *India*—C. R. Rao. *Italy*— P. V. Sukhatme. *Japan*—T. Kitagawa. *Mexico*—Ana Maria Flores, J. Nieto de Pascual. *Portugal*—Flavio Resende. *Switzerland*—A. Linder. *United States of America*—C. A. Bicking, C. I. Bliss, A. E. Brandt, W. G. Cochran, G. M. Cox, B. B. Day, E. R. Dempster, Th. Dobzansky, M. H. Hansen, P. G. Homeyer, E. Lukacs, G. O. Mott, W. R. Pabst, W. J. Youden.

BRIEF OF PRESIDENTIAL ADDRESS:

THE 1954 TRIAL OF THE POLIOMYELITIS VACCINE IN THE UNITED STATES

William G. Cochran

This trial represents an important application of biometrical principles in the struggle against disease. The experimental subjects were children in the first three grades or classes of school, of ages about 6–9 years. In terms of numbers of subjects the experiment may be the largest that has ever been conducted.

MAJOR DIFFICULTIES IN THE CONDUCT OF A TRIAL

(1) Poliomyelitis is a relatively rare disease. From past experience, the rate of paralytic polio in the study areas might be anticipated to be about 30 cases per 100,000 children aged 6–9 years. Given this attack rate, table I shows the probability of obtaining a statistically significant result (5% level) for various numbers of children and for various degrees of *true* effectiveness of the vaccine. With a vaccine that actually was 50% effective, about half a million children would be needed to make the risk of an inconclusive result small. Table II shows

TABLE I

Probability of obtaining a significant result (5% level)

No. of children in trial	True effectiveness of vaccine		
	50%	70%	90%
200,000	0.59	0.91	>0.99
400,000	0.88	>0.99	>0.99
600,000	0.97	>0.99	>0.99

TABLE II

Confidence limits for the true effectiveness

No. of children in trial	Observed effectiveness of vaccine		
	50%	70%	90%
200,000	2%—75%	34%—88%	68%—98%
400,000	20%—69%	48%—83%	77%—96%
600,000	27%—66%	53%—81%	80%—95%

the 95% confidence limits that would be obtained for the true effectiveness, if the observed effectiveness in the trial turned out to be 50%, or 70%, or 90%. Even with 600,000 children the true effectiveness can be none too well determined, except for a vaccine with an effectiveness up in the 90% range.

(2) The disease is difficult to diagnose: even in the paralytic form mistakes can be made. Some of the indefiniteness can be removed by adopting stringent criteria for the definition of a case. However, this device, if carried too far, may defeat its own ends by reducing the "accepted" cases to a very small number.

(3) The vaccination itself required 3 injections, the second given one week and the third 5 weeks after the first.

(4) The experiment subjects were children. Would parents give permission? Would physicians, health officers and medical societies give and encourage cooperation?

(5) Some biometricians have learned from bitter experience to take a pessimistic view of the prospects of success of *any* large trial with human subjects. Procedures that are essential for valid comparisons are apt to be cast aside as administratively impractical: instructions issued from a central office may be misread, misinterpreted or simply changed by persons a long way off; incomplete record forms and missing data flourish, and so on.

THE PLAN OF THE STUDY

The National Foundation for Infantile Paralysis invited the states individually to participate in the trial. If a state agreed, the vaccine was tested in all schools in certain counties within the state that had been selected by the Foundation. In order that the evaluation of the vaccine should be independent of the Foundation, the operation of the trial and the analysis of results were placed under the direction of Dr. Thomas Francis, with headquarters at the University of Michigan.

The plan announced by the Foundation was that the second-grade children in a participating school would receive the vaccine, while first and third grade children would remain unvaccinated to serve as controls.

This plan is subject to a number of potential biases. It requires the assumption that the attack rate among second-grade children is the same as the average attack rate amongst first and third grade children. Secondly, not all parents of second-grade children would allow their children to be vaccinated. Actually, 69% of them gave permission. Thus the plan compares a selected 69% of the second-grade children with the other two grades. There are epidemiological grounds for

arguing that this selection biases the results against the vaccine. Further, in any suspected case of the disease, it would be easy to discover whether a child had been vaccinated. This fact could create an unintentional bias in diagnosis by the local physician and could affect the completeness of reporting, as well as the precautions taken by parents for their children in the event of an epidemic.

It might be argued that the cumulative effect of these sources of biases was bound to be small and that results could not be seriously distorted if the vaccine was potent. But this assertion cannot be proved and with this method there must remain an element of doubt.

This plan was followed in 33 states, with 222,000 second-grade children vaccinated and 725,000 controls from the first and third grades.

A number of states adopted a different plan. Participating children in the 3 grades were divided at random into two groups. One group received three shots of the vaccine: the other received three shots of an inert fluid made up to have the same appearance as the shots of vaccine. The two treatments were distinguished by code numbers accessible only to those in charge of the study.

This plan raised more administrative difficulties than the first plan, but was free from the sources of biases that have been mentioned with respect to the first plan. All diagnoses, reporting and classification of cases, and all except the final stages of the analysis were done in ignorance of whether the child had received vaccine or placebo.

This plan was adopted in 11 states. Each treatment (vaccine or placebo) was represented by some 201,000 children. It is highly encouraging to biometricians that state officers and epidemiologists in these states expressed their preference for this plan, despite its many difficulties of execution.

Space permits mention of only a few aspects of the operation of the experiment. Collection of data was a formidable task, involving large numbers of letters, telegrams, telephone calls, regional and local conferences and special visits by members of the evaluation team to local areas. These efforts produced a high degree of completeness: missing data were of negligible importance.

Diagnoses were obtained in the following manner. When a suspected case appeared, a clinical history, including spinal fluid examination and blood and stool specimens, was made by the local physician on a standard form. A muscle examination was conducted by a physical therapist 10–20 days after onset, and a further examination 50–70 days after onset: each muscle report was reviewed by a local physician experienced in the clinical aspects of polio.

On the basis of these local records, a team of experts recruited by

the evaluation center at Michigan classified each case into one of the categories: (1) not polio (2) suspect (3) non-paralytic polio and (4) paralytic polio. The paralytic cases were further classified as to type and severity of paralysis. All these diagnoses were made by criteria that had been thrashed out and written down in advance by the team.

For record keeping and statistical analysis at the evaluation center itself, a small team of persons familiar with the handling and processing of large masses of data was obtained on leave of absence from the Bureau of the Census.

SOME RESULTS

Results were analysed and presented separately for the two plans. Areas covered by the original plan were called *observed* areas, while those that participated in the second plan were called *placebo* areas.

Table III shows the numbers of cases and the case rates per 100,000 children in the two areas. Incidentally, the paralytic case rates among non-vaccinated children were 43 in the placebo area and 44 in the

TABLE III
Cases and case rates per 100,000 children

Areas	No. of children in study	Polio cases			
		Paralytic		Non-paralytic	
		No.	Rate	No.	Rate
Placebo					
Vaccinated	200,745	33	16	24	12
Placebo	201,229	115	57	27	13
Not inoculated	338,778	121	36	36	11
Observed					
Vaccinated	221,998	38	17	18	8
Controls	725,173	330	46	61	8
2nd Grade not inoculated	123,605	43	35	11	9

observed area. Both rates were substantially above the anticipated rate of 30 which I used in discussing the needed sample size, so that the study had good fortune in not taking place during a year of unduly low incidence. The cases included in the results were all those that occurred between two weeks after the third injection and December 31, 1954.

In the placebo areas, paralytic case rates were 16 for vaccinated children and 57 for unvaccinated children. This gives an estimated effectiveness of 72%. In the observed areas the corresponding rates were 17 and 46, with an indicated effectiveness of 64%.

For non-paralytic cases, the rates were practically the same in vaccinated and control groups in both the placebo and observed areas. Although this result is somewhat unexpected, at least to a layman, it need not give concern from a public health point of view, since non-paralytic polio is not a major hazard like the paralytic form of the disease.

Table III also carries two lines marked "Not inoculated." In placebo areas this line refers to children in all three grades whose parents did not give permission to participate, plus a small number of children who received only one or two shots of placebo. In observed areas this group comprizes second-grade children whose parents did not request participation. In both areas the "not inoculated" group showed lower paralytic rates than the corresponding controls (36 against 57 and 35 against 46).

A difference in this direction had been anticipated on epidemiological grounds. Children of parents who withheld permission might be expected to be of a somewhat lower economic level than participating children, and to have acquired a greater degree of natural protection against polio through a previous subclinical attack of the disease. This type of selective bias has no effect on the results in the placebo areas, in which the comparison between vaccine and placebo was made entirely from participating children. In the observed areas, the bias would tend to reduce the apparent effectiveness of the vaccine. The fact that the vaccine showed lower effectiveness in the observed than in the placebo areas (64% against 72%) is in line with this explanation. A special sample survey that was made of participating and non-participating parents also tended to confirm the presence of a difference in economic level.

Table IV shows the estimated effectiveness of the vaccine as obtained from two more stringent criteria of classification. The main points to note are that the more severe criteria bring about some increase in the estimated effectiveness, and that the effectiveness figures run consistently about 10% lower in the observed than in the placebo areas.

The problem of making tests of significance and constructing confidence limits requires some consideration. One approach is to assume that the number of cases under a specific treatment in a school will follow a Poisson distribution. The total number of cases over all schools will then also follow a Poisson distribution, and the tests and

TABLE IV

Results given by more severe diagnostic criteria

| Diagnosis | Placebo areas | | | 95% limits for effectiveness |
| | No. of cases | | Estimated effectiveness % | |
	Vac.	Control		
Paralytic	33	115	72	57—81
Lab. confirmed	10	68	85	71—93
Positive virus obtained	15	70	80	62—89
	Observed areas			
Paralytic	38	330	62	47—74
Lab. confirmed	16	198	74	56—86
Positive virus obtained	20	210	69	50—82

limits can be constructed from Poisson theory. A more conservative approach, which avoids the Poisson assumption, is to regard the county as the basic sampling unit. The tests and limits are made by "continuous variable" theory, using the interaction with counties as the measure of error.

By either approach there is no doubt of the statistical significance of the beneficial effect of vaccine on paralytic cases. Confidence limits obtained by the Poisson approach appear in table IV, and serve to indicate the realm of uncertainty in our information as to the real effectiveness of the vaccine. The corresponding limits as obtained from the continuous variable approach would be somewhat wider.

Much credit is due to all who cooperated in this trial, and particularly to Dr. Francis and his staff, for the high standards maintained throughout the operation, despite the huge numbers of children to be processed. Among the many factors that contributed to give a fully valid comparison in the placebo areas, some of the most important were: (1) Randomization of children between vaccine and placebo (2) Keeping those concerned with case finding, diagnosis and classification in ignorance as to the treatment given to any child (3) Adoption of detailed criteria for the final diagnosis and classification and (4) Willingness to take endless pains to secure completeness and uniformity in reporting.

The question of the safety of the vaccine when given to such large numbers was of great concern. Special reports on all deaths of children,

from whatever cause, records of unusual reactions following shots, and studies of absenteeism from schools following shots were made. None of these indicated any basis for apprehension about the safety of the vaccine in this trial.

No discussion has been given here of a large volume of laboratory work designed to test the lots of vaccine, to study the rises in antibody levels following vaccination and to attempt to identify the virus from any case.

The Summary Report issued by the Vaccine Evaluation Center, University of Michigan, from which the data presented here were taken, should be consulted for a much more adequate account of the trial.

ABSTRACTS OF PAPERS

International Biometric Symposium, Instituto de Educacao Carlos Gomes, Campinas, Brazil, July 4–9, 1955

343 R. A. FISHER. **The Contribution of Biometry to Plant Breeding.**

The lecturer listed these contributions under the headings of

(1) Experimental Design
(2) Biometrical Genetics, in the sense of K. Mather
(3) Biometrical applications to classical Mendelian genetics.

He emphasised that the art of plant improvement needed in addition to genetical knowledge, the role of which will doubtless increase as greater refinement and penetration is attempted, a basic familiarity with agricultural science, and especially with the art of carrying out field trials with accuracy. Indeed the greater part of the money value due to plant improvement to date must be ascribed to improvement in field plot techniques.

After reviewing some of the concepts of biometrical genetics in the analysis of variance of the metrical values in a plant population, the lecturer turned to the increased complexity introduced into classical genetics by the study of polysomic inheritance, and the extended use in this field of complex analyses depending on observed frequencies.

344 EVERETT R. DEMPSTER. **Genetic Models in Relation to Animal Breeding.**

As gains from selective breeding diminish, attention is necessarily shifted from methods for obtaining the most improvement in a single generation to methods for achieving maximum gains over a span of many generations. On the basis of simplified assumptions, the ultimate gain from indefinitely repeated mass selections would lie between $2Nz\sigma h^2$ and $2Nz\sigma \left[1 + \dfrac{2n - 1}{2n} \right]$, where N and n are the numbers in the population and of parents respectively, z is the height of the ordinate separating selected parents from the remainder of the population, and h^2 is the heritability. This is maximized when half, or very slightly more than half, of the population is used as parents in each generation,

and hence when selection intensity is relatively low. In moderate to large sized populations, however, this formulation would apply only to those loci where the differential effects of alleles are small to exceedingly small respectively. This demonstrates that quantitative predictions require, among other information, a knowledge of the distribution of the magnitudes of differential allelic effects at different loci. Even a qualitative conclusion—that intense selection will reduce ultimate gains —would be justified only if (as may be true) other deviations from the simplified model, such as non-additive genetic variance, linkage, and negative genetic correlation between natural fitness and characters selected for, also tend to produce a similar relationship.

It is clear that useful predictions regarding long term gains require much more knowledge than is currently available regarding variation in breeding populations. The difficulties of obtaining such information are so great that no reasonable method of attack can be neglected. One method involves deductions from what is known or may be reasonably assumed in regard to mutation rates, natural selection, and Mendelian inheritance. Such deductions have been made on the basis of simple assumptions, but only recently has much attention been paid to selection pressures variable in space and time. It is shown by an example that, under some circumstances, alleles otherwise subject to elimination by natural selection or drift could be retained in a population indefinitely if the selection pressures of given average values, were variable in time instead of steady.

345 F. G. BRIEGER. **Behavior of Autogamic Populations and Heterotic Genes.**

The requirements of applied genetics are undergoing a very significant change. In large regions of the world, mainly outside Europe and at least parts of the USA, the climatic, edaphic and economical requirements are extremely diversified and in excess to the number of people engaged in breeding work. Thus only such a degree of homogeneity seems to be desirable which is still compatible with a sufficient amount of plasticity, in order that the improved material may serve over larger areas. To achieve this, a shift of methods is necessary, giving preference to breeding methods which deal not with well defined pedigree lines, but with populations. In order to plan efficient work in population breeding, it is necessary to work out models of the genetic constitution of populations, taking into consideration the effects of recurrent mutation and selective activities, which may be applied to both panmictic and autogamous populations and to all intermediate cases which may occur.

These formulae should give the frequencies for the three basic genotypes (AA, AA', and $A'A'$).

Using a basic approach given by R. A. Fisher, we may start from equal numbers of individuals of the three genotypes, and determine how many individuals or gametes of each are left after all processes of selection have played their part. If these numbers be a, b and c, we may determine the coefficients of survival of the two homozygotes with reference to the heterozygotes by dividing by the remaining frequency b of these heterozygotes. Thus the survival value of heterozygotes is unity, and we obtain two coefficients only: $R_A = a/b$ for the homozygotes AA and $R_{A'} = c/b$, for the homozygotes $A'A'$, which may have any value from zero (complete elimination of homozygotes) to infinite (complete elimination of heterozygotes). We may then determine the triple proportions for a manofactorial population at equilibrium and for both panmictic and autogamous populations, with reference to the three most important types of mutant genes: recessive subviables, neutrals, and heterotics:

| | Panmictic | | |
	AA	AA'	$A'A'$
Neutrals	v^2 :	$2uv$	: u^2
Rec. sub-viables:	1 :	$2\sqrt{\dfrac{\mu}{1 - R_{A'}}}$	$: \dfrac{\mu}{1 - R_{A'}}$
Heterotics:	$(1 - R_{A'})^2 : 2(1 - R_A) \cdot (1 - R_{A'}) : (1 - R_A)^2$		

| | Autogamous | | |
	AA	AA'	$A'A'$
Neutrals	v :	$8uv$	: u
Rec. sub-viables:	1 :	$4u$	$: \dfrac{u}{1 - R_{A'}}$
Heterotics:	$(1 - 2R_{A'}) : 4(1 - 2R_A) \cdot (1 - 2R_{A'}) : (1 - 2R_A)$		

In panmictic populations all three neutral or modifier genotypes are of the same order, all being sterms of second order. In autogamous populations however the frequencies of heterozygotes are proportionally smaller than those of homozygotes, and the population will actually consist of a mixture of "pure lines". The frequency of subviable

homozygotes is small, but equal in both types of populations. The maximum frequency of heterozygotes in autogamous populations is very small and equal to four times the mutation rate only, since in each generation there appear by mutation 2μ heterozygotes, while at the same time half of all heterozygotes are lost by segregation. Heterotic loci must be very rare or even practically non-existent in self-fertilized populations, since only such loci behave as heterotics which contain two alleles, with homozygotes of a viability less than half of heterozygotes (R_A and R'_A both smaller than 0.5).

We may also use these formulae to explain the situation of special cases, such as that of heterosis in panmictic species. Thus it can be shown easily that about 1.000 loci of subviable recessives give results comparable to about 50 heterotic loci.

We may obtain also models for multifactorial segregations, by calculating the terms of the respective trinomials, with an exponent n equal to the number of loci involved. The frequencies of special cases of gene interaction may be calculated by uniting the respective frequencies of classes before termination of the calculation of the polynomial terms.

Finally we may easily determine the loss of a population caused by selection. The above formulae refer to the situation at the beginning of a generation, and by multiplying the frequencies of homozygotes by total or partial survival values and taking the differences from the original frequencies, the percentage loss may be determined.

346   S. C. PEARCE. **The Specific Problems of Experimental Design and Technique in Perennial Crops.**

Problems are considered in three classes, (1) those in scientific approach arising from the small number of experiments possible within a limited period, (2) mathematical problems in design and interpretation, and (3) those concerning the variability and measurement of plants.

Experimentation with perennial plants is laborious, and survey methods difficult of application on account of the many factors to be disentangled. A further possibility is to use sequential methods to investigate the opinions of those with the experience to give a useful judgement on a specific question, but it is still essential to [make the best use of the limited number of trials that can be done. This involves] see[ing] each experiment against the corpus of existing knowledge, [and not merely accumulating facts which cannot be explained, as is often done profitably with annuals.] The experimenter should [in fact, think about the subject, and then] use his experiment to confirm or

refute his chain of thought, [and not merely to establish phenomena in isolation. If this is the aim,] it is not enough to measure only crop [or whatever is under study;] records are required of anything that may lead to understanding of the cropping situation.

Three problems in the second group were mentioned: (a) The evolution of row and column designs, for these often enable outside trees as well as inside ones to be used in an experiment, and also permit the addition of treatments to a trial in randomized blocks. (b) The consequences of using parts of an organism, e.g., the branches of a tree, as the plots of an experiment, bearing in mind the possibility of the treatments applied to one plot affecting also the other plots of the block. (c) The combining of several years' results, a problem that awaits satisfactory progress, unlike the other two in which useful advantages have been made.

In the third group, study needs to be made of the relative importance of the various sources of variation to avoid waste of effort in controlling those of less effect. Secondly, and almost certainly the dominating problem still to be solved, is the measurement of important characters in the living tree; only when the few characters that can now be measured are supplemented will it be possible to understand how treatments have their effect, and thus to make best use of the experiments possible.

347 A. CONAGIN AND C. G. FRAGA. **Design and Analysis of Coffee Experiments.**

An outline is given of the designs and procedures of statistical analysis in coffee experiments at the Instituto Agronomico de Campinas. Four groups of experiments were considered: 1) fertilizer tests; 2) varietal trials; 3) progeny tests; and 4) miscellaneous experiments.

Under 1) the writers discussed old experiments with systematic layouts and more recent ones with randomized designs. A factorial experiment supplied conclusive results within a few years. The problem of changing some of the treatments arose in one of the experiments and the solution proposed by the writers was discussed.

Under 2) the treatment of data used by W. L. Stevens (1949) in the analysis of a systematic experiment comparing coffee varieties was commented in detail.

The evolution in designs used for progeny comparisons was discussed under 3). In early tests progenies were compared on the basis of rows of 20 plants without replications; later the designs were changed to replicated plots of 4 plants per plot, and more recently to plots of a single plant with a higher number of replications.

The results of spacing trials and of tests comparing several methods of planting coffee were discussed under 4). The plot size for coffee experiment was also discussed, based on individual yield records from a planting of the Bourbon variety.

348 F. PIMENTEL GOMES. **Methods of Describing Crop Response to Fertilizers in Perennial Crops.**

The author discusses briefly the advantages and inconveniences of fitting polynomials and Mitscherlich's equation to data corresponding to graded levels of fertilizers, the conditions which the data are supposed to fulfill in order for the fitting of Mitscherlich's law to be possible, spacing of levels to increase the probability of a good fitting, and further advisable procedure to ensure sound experimentation having in view the fitting of Mitscherlich's law.

For the special case of sugar cane, one may try to describe crop response to fertilization in a whole cycle or in each harvest separately. Factorial experiments are usually to be preferred, but in some instances just one nutrient is varied in the experiment. Such was the case in an experiment carried out in the Usina Monte Alegre (Piracicaba), by E. M. Cardoso, with the levels 0, 10, 20, 30, 40 and 50 kg/ha of K_2O.

Mitscherlich's equation was $y = 98.2 [1 - 10^{-2.03(x+0.416)}]$ $t./ha.$ The most profitable level of fertilization was $x^* = 58$ $kg./ha.$ of K_2O. For data obtained by Strauss, in Pernambuco, in twenty-three $3 \times 3 \times 3$ factorial NPK experiments with sugar cane, the equations and most profitable levels were for Phosphorus: $y = 73.32 [1 - 10^{-0.918(x+0.490)}]$ $t./ha.$, $x^* = 82$ $kg.$ of P_2O_5 per hectare; for Nitrogen: $y = 71.97$ $[1 - 10^{-0.366(x+1.789)}]$ $t./ha.$, $x^* = -39$ $kg.$ of N per hectare; for Potash: $y = 65.49 [1 - 10^{-0.984(x+0.871)}]$ $t./ha.$, $x^* = 68$ $kg.$ of K_2O per hectare.

It is usual in Brazil to plant sugar cane in such a way that it has a 3 1/2 year cycle, with three harvests. The first ratoon yields around 70%, and the second around 50% of the first harvest, which is, therefore, the most important one. So, a way to solve the problem of determining the most profitable level of fertilization for each crop of the sugar cane cycle could be to obtain firstly the most profitable level for the first harvest and then, taking this for granted, study separately how much fertilizer should be used in the first ratoon. Again, assuming that the most profitable levels for the first and second harvests are used, we should try to find out what is the best level for the second ratoon. However, since in most cases fertilizers are used only in the first harvest, another way to describe sugar cane response to fertilizers would be to take together the first harvest and the two ratoons. This was done in

the analysis of a $2 \times 4 \times 3$ factorial experiment with NPK, carried out at Usina Itaiquara by the Sugar Cane Department of the Instituto Agronomico.

The levels of P_2O_5 were 0, 60, 90 and 120 $kg./ha.$ Equation obtained was $y = 316.50 [1 - 10^{-0.859(x+1.23)}]$ $t./ha.$ and the most profitable level was $x^* = 107$ $kg.$ of P_2O_5 per hectare.

A further way of analysing factorial experiments with fertilizers, either for annual crops or perennials would be the fitting of a Taylor series with several independent variables, following the sequential methods of Box and Wilson. However, the author thinks that such a way is not advisable because: 1) sequential methods are too slow in agricultural research; 2) the asymptotic regression given by Mitscherlich's equation seems to be suitable in most cases, which is not true with respect to polynomials; 3) approximating polynomials obtained from Taylor's series are good only to describe local properties of a curve, and this is not enough when dealing with prices, which, when changing, shift our attention to other portions of the curve. In agriculture a similar shifting may be caused also by the introduction of new varieties or by other means of increasing the yield.

349 GERTRUDE M. COX. **Some Recent Advances in Experimental Designs, with Particular Reference to Estimating Response Surfaces.**

A brief survey of the basic, or classic, designs used in experimentation was given. The second section of the paper presented variation in these basic designs along with designs currently being developed. Those discussed were (1) balanced groups with covariance (2) change-over trials (3) doubly balanced incomplete block designs (4) partially balanced incomplete block designs with two associates (5) chain block designs and (6) paired comparisons.

The third and major portion of the paper dealt with the designs being used to secure an estimate of the optimal point and to explore the nature of the response surface in the vicinity of this optimum. Examples were given illustrating the use of composite and rotatable designs in actual experiments.

350 W. J. YOUDEN. (National Bureau of Standards, Washington, D. C.). **Design of Experiments in the Physical Sciences.**

Experiments in the physical sciences reveal certain features not usually exhibited by classical designs used in agricultural field trials.

(1) Block size is often uniquely determined by the apparatus, or physical equipment, or material investigated. (2) Block effects are almost always of interest and are sometimes of primary interest. Blocks may be instruments, or positions in the equipment, and these are enduring entities. (3) The relatively high precision of physical measurements means that little replication is needed and designs involving considerable replication are unacceptable. (4) Sometimes it is highly desirable not to specify all the treatments in advance because the experimenter desires to obtain some results, say at a few temperatures, before specifying the other levels. This leads to the use of designs which may be incorporated in a larger design when the first results are in hand. (5) Measurements are usually obtained in a time sequence and provision for instrumental or environmental drift is frequently necessary. (6) Finally the experimental situation is sometimes of a unique nature. For example, the environmental temperature instead of being held constant may, for good reason, be made to rise steadily. Special designs allow more precisely for the temperature effect than simple block arrangements.

351 GERALDO LEME DA ROCHA. **Grazing Experiments in the State of São Paulo.**

Methods of pasture management are of great importance for the dairy and beef cattle industry in the state of São Paulo. A number of experiments are being conducted in various parts of the state by the author and collaborators. An outline of the experiments concerning rotational grazing is being given here.

I. *Experiment in the Paraíba River Valley Region*

Four species of grasses, (Gordura grass—*Melinis Minutiflora Beauv.*; Jesuita grass—*Axonopus compressus* Beauv.; Colonião de Tanganika grass—*Panicum maximum* Jacq.; Sempre Verde grass—*Panicum maximum* var. *gongyloides* Jacq.), were compared in grazing tests using yearling dairy cattle. Thirty-two paddocks of 5,000 sq. m. each were utilized. Although no replication was made, the experiment was divided in 4 blocks of 8 paddocks each. Within the blocks the treatments were randomized. Each species of grass was planted with and without fertilizers.

Three animals in each of the 8 paddocks in block *I* started grazing on the same day and were rotated according to the condition of the sward and liveweight variations. Weighing was done once a week. Data obtained after 9 months showed no correspondence between

liveweight variation and subjective judgment of the sward. From then on the animals were left grazing in each paddock for 8—14 days and were then moved over to the next paddock (any surplus grass was grazed by "followers"). Botanical analyses were made twice a year (December-January and July-August).

II. *Experiment near Ribeirão Prêto (red soil)*

The same layout of the experiment described above was followed. One hectare paddocks of the following species of grasses were compared: Colonião grass—*Panicum maximum* Jacq.; Makari-Kari grass—*P. coloratum* L.; Gordura grass, and Jaraguá grass—*Hyparrhenia rufa* (Nees) Stapf. Grazing in this case was done with groups of 10 yearling beef cattle.

III. *Nova Odessa Experiment*

The experiment in this area aimed at finding out the best management for gordura grass swards on the basis of *use* and *rest*. Groups of 6 animals were put to graze in 5,000 sq. m. paddocks, according to the following schedule:

> Treatment A—4 days use—20 days rest
> B—6 days use—30 days rest
> C—8 days use—40 days rest
> D—continuous grazing

Botanical analyses were carried out as described before.

IV. *Collina Stud Farm Experiment*

This experiment was similar to III, but had an improved design. Sixteen 5,000 sq. m. paddocks were available. Four replications of each of 4 treatments were compared:

> Treatment A—4 days use—12 days rest
> B—6 days use—18 days rest
> C—8 days use—24 days rest
> D—continuous grazing

Grazing was done with mares. During the resting period the animals grazed on the common swards of the farm. Botanical analyses were made as before.

V. Another experiment is being carried out in the Paraiba River Valley Region. Kikuiu grass—*Pennisetum clandestinum* Hochst ex Chiov.,

and Rhodes grass—*Chloris gayana* Kunth, are being tested in two different ways: rotational grazing versus strip grazing. This comparison is based on experiments designed by Holmes and carried out at the Dairy Hannah Institute, Ayr, Scotland.

352 G. O. MOTT. (Purdue University, Lafayette, Indiana). **The Grazing Trial for Measuring the Output of Pasture.**

The objective of the grazing trial is to measure the quality of herbage produced by a pasture and the yield of animal product per unit area. It yields information useful to both the Agronomist and Animal Husbandman in that the output per animal is an indication of the quality of forage, the carrying capacity in terms of animal days is a reliable index of the herbage production per unit area, and the livestock product per acre is an indication of both the quality and quantity. The daily output per animal is a function of the nutritive value of the forage, the rate of intake and the physiological characteristics of the animal. The performance of the animal is also greatly affected by the grazing pressure and the opportunity for selective grazing.

The size and sources of the experimental errors associated with the grazing trial differ for the several units of measure. Both the pasture variability and that which can be attributed to the animal has to be considered. A study of the sizes of the errors for the various units of measure points to the need for at least three and preferably five or more field replications in a grazing trial. The size of the field for each replication should be sufficient to supply at least two animals with adequate herbage.

The source of bias most commonly encountered in the grazing trial is the failure of the investigator to estimate carrying capacity at the optimum. If the pasture is overgrazed, the number of animal days will be overestimated, the daily performance of the animal will be less than that expected at the optimum and the product per acre will also be underestimated. If on the other hand the pasture is not grazed to capacity, then the number of animal days will be low, the daily gains may be slightly overestimated and the product per acre will be underestimated due to the failure to utilize the forage produced.

Simple designs are usually indicated for the grazing trial due to the limited number of treatments and replications involved. Techniques used to reduce the errors due to the animal in the feeding trial are also useful in the grazing trials for variates such as previous performance and initial weights.

MORRIS H. HANSEN AND JOSEPH STEINBERG. (U. S. **353** Bureau of the Census, Washington, D. C.). **Control of Errors in Surveys.**

In the evaluation of the Current Population Survey—(a monthly population survey made by the U.S. Bureau of the Census to estimate labor-force characteristics and other population data)—control of errors has been sought through the traditional devices of selection, training and supervision of personnel. Efforts to increase the objectivity of the measurement of error in the enumeration and interview processes are pursued by a formal quality control procedure based on re-interviews.

During a twelve-month period the recheck of coverage indicated minimal errors by a preponderance of the enumerators (85% of enumerators had zero error rate) and concentration of errors within a small portion of the enumerators (5% of enumerators were the source of approximately three-quarters of the errors); however, the recheck of information obtained during the interview indicates that approximately 60% of the enumerators have varying amounts of differences between the original and recheck results. Various possible causes of discrepancies in content were examined. The respondent appears to contribute as much or more heavily to the differences than does the interviewer. The sources of error are difficult to identify and consequently the check results are difficult to interpret in controlling the work of individual interviewers. Examination of net differences between original interviews and reinterviews in terms of estimated standard errors indicates that for the most part the interviewing in the Current Population Survey can be considered to be under control.

Experimentation is continuing on various sources of error and the control of observational errors.

354 P. V. SUKHATME, (FAO) and V. G. PANSE, (ICAR). **Sampling Technique for Estimating the Catch of Sea Fish.**

The paper describes a sampling method developed in India for estimating the monthly catch brought to the coast by fishing boats. It is divided into two parts—the first dealing with the sampling procedure for estimating the daily catch at selected sections of the coast and the second dealing with the choice of optimum first-stage unit for sampling and the number of such units for estimating the monthly catch for the entire coast with given precision.

The sampling procedure for estimating the daily catch is based on the study of hourly landings at a number of selected sections along the

coast, and it is concluded that a systematic selection involving two visits of three hours (or two, in case the journey time to the coast is shorter) during a day is both a practical and efficient scheme of sampling at selected sections. The section of the coast found most suitable for use as sampling unit for observation is a landing centre. The paper then gives the result of a study to determine the optimum number of days for which a selected centre should be observed in succession, based on the data collected at 61 landing centres for two months and concludes that a centre × day is the optimum unit of observation. The number of centres to be selected daily for estimating the monthly catch for the entire coast with 5 per cent error is placed between 15 and 22.

Finally, a brief description is given of the surveys conducted in India in the course of which the above technique was developed. The present surveys cover a coast of nearly 500 miles in length. It is proposed to extend the surveys to cover the entire coast of India as the normal method of estimating the monthly catch of marine fish.

355 ENRIQUE CANSADO. **Sampling Without Replacement from Finite Populations.**

Although Hansen and Hurwitz (1953), Midzuno (1950), Narain (1951), Horwitz and Thompson (1952) and Yates and Grundy (1953) made important contributions to a general theory of sampling without replacement from finite populations, the presentation of this general theory did not satisfy the standards with respect to rigour and systematicity, that are now prevalent in the expositions of other branches of Calculus of Probability and Mathematical Statistics.

In this paper this well-known theory is presented in a form which, it is hoped, will reduce some of the aforementioned deficiencies. Starting with the fundamental set of selecting probabilities $P_1(u_i)$, $P_2(u_i/u_j)$ $\cdots$, $P_n(u_i/u_{h_{n-1}} , \cdots , u_{h_1})$, which defines completely the sampling scheme considered, formulae are given for the probabilities $P_1(u_i)$, $P_2(u_i)$, $\cdots$, $P_n(u_i)$ of selecting the unit u_i at the first, second, $\cdots$, n-th draw. From these are obtained the probabilities $P(u_i)$ of inclusion of the unit u_i in a sample of size n. Formulae are also considered which give, from the fundamental set of selection probabilities, the probabilities $P_{rs}(u_iu_j)$ of selecting the unit u_i at the r-th draw and the unit u_j at the s-th draw. From these are obtained the probabilities $P(u_iu_j)$ of including both the units u_i and u_j in a sample of size n.

On this basis it was easy to obtain the formulae for the mathematical expectations of the sums and the product-sums of the observed values in

a sample of size n. Then $\hat{T} = \sum_{r=1}^{n} \frac{x_r}{P(u_r)}$ was considered as an estimator of the population total $T = \sum_{i=1}^{N} X_i$. It was, then, readily shown that $\hat{T}$ is unbiased and formulae are obtained for the sampling variance $V(\hat{T})$ of this estimator.

The paper ends with a consideration of two unbiased estimators of this sampling variance:

$$\hat{V}_1(\hat{T}) = \sum_{i=1}^{n} \frac{1 - P(n_i)}{P^2(u_i)} x_i + \sum_{i \neq j = 1}^{n} \frac{P(u_i u_j) - P(u_i)P(u_j)}{P(u_i)P(u_j)P(u_i u_j)} x_i x_j$$

given by Narain and Horwitz and Thompson, and

$$V_2(\hat{T}) = \frac{1}{2} \sum_{i \neq j = 1}^{n} \frac{P(u_i)P(u_j) - P(u_i u_j)}{P(u_i u_j)} \left\{ \frac{x_i}{P(u_i)} - \frac{x_i}{P(u_j)} \right\}^2$$

given by Yates and Grundy.

356 C. I. BLISS. **Confidence Limits for Measuring the Precision of Bioassays.**

In measuring the precision of an assayed potency, confidence or fiducial limits have the advantage over the standard error of taking full account of the size of the assay and the precision of its slope. Experimentally, bioassays can be divided into two types, (1) those based upon the mean threshold dose, measured directly in each test animal, and (2) those where the size of the reaction at selected dosage levels is the dependent variable and relative potency must be inferred by converting the response back to units of dose or log-dose. For the few assays of the first type, the confidence interval of the log-relative potency M' is that for a mean difference or for the difference between two means, both well known and simple calculations. For the much larger group of assays comprising the second type, potency depends upon the ratio of two statistics. If the dosage-response curve is linear with arithmetic dosage units, potency is computed from the ratio of two slopes, one for the Standard preparation and the other for the Unknown. If instead the response plots linearly against the log-dose, the more common type, log-potency is computed from the ratio of a difference in the mean response $(\bar{y}_U - \bar{y}_S)$ to a slope (b).

One of the simplest forms for computing the confidence limits for a ratio was that introduced in 1944 by Marks for balanced cross-over assays for insulin. His equation has been generalized for all assays based upon the ratio of two statistics with little loss in its inherent

simplicity. If the log-relative potency is computed as $M' = (\bar{y}_U - \bar{y}_S)/b = a/b$ and the assay is balanced so that the numerator and the denominator are independent, the confidence limits for M' can be determined as

$$X_{M'} = CM' \pm \sqrt{(C - 1)(CM'^2 + v_{aa}/v_{bb})}$$

where $C = b^2/(b^2 - v_{bb}t^2)$, and the variances of a (v_{aa}) and of b (v_{bb}) have a ratio depending upon the design of the assay and is determinable in advance. Factorial log-ratio assays follow this pattern, including assays in balanced pairs and with more than one Unknown. It is also applicable to all-or-none assays and to assays based upon the ratio of two mean threshold doses.

When the numerator and denominator of M' are not known to be independent, their covariance leads to an additional term both inside and outside the radical in the above equation. The effect of this co-variance in assays arranged in randomized blocks or groups, in assays needing replacements, in assays with a single test animal having a changing sensitivity, and in slope-ratio assays are considered. Illustrative numerical examples have been drawn in part from recent studies connected with the new U.S.P. XV.

357 D. J. FINNEY. (The University, Aberdeen, Scotland). **Cross-Over and Single-Subject Designs for 4-Point Assays.**

Many biological assays can be increased in precision by measuring responses to different doses successively on the same subjects. Adoption of different dose sequences for different subjects produces cross-over designs. The statistical analysis of the results of these involves some consideration of time series. In particular, the possibilities of correlation between components of residual error, of residual effects of past doses, and of autoregressive influences of one response on its successors need to be considered. This paper presented five different models that may be appropriate to bioassays.

Using these models, the analysis of 4-point parallel line assays for various assay designs (twin cross-over and Latin square types) was then discussed. Lucas's findings that the existence of residual influences need not bias these designs in respect of treatment comparisons were confirmed, and the special procedures for estimating error variances that Patterson suggested were developed. New features peculiar to bioassay, notably the presence of several independent or semi-independent validity tests and estimates of relative potency, were considered in detail.

A new class of designs for single-subject assays, possessing a property of serial balance, was described, and the statistical analysis of these was explained.

358 O. G. BRIER, M. SIQUEIRA AND P. M. FREIRE. Application of Bioassay Methods to Complement Fixation.

Quantitative complement fixation tests were performed with varying dilutions of sera from syphilitic patients and a constant optimal amount of cardiolipin antigen. A large, constant amount of complement was available in the fixation mixtures, and from the spectrophotometric titration of the residual hemolytic activity, the numbers of 50 per cent hemolytic units of complement bound to the various amounts of antibody were estimated.

When the dose of antibody-containing serum was expressed in the logarithmic scale, over a certain range there was a linear relation with the number of complement units fixed by any constant volume of the mixture. The adequacy of this linear relation was established in replicated tests with 4 syphilitic sera. As tested by analyses of variance, the data for each serum agreed with the corresponding regression lines within the limits of chance variation.

The reproducibility of assay results and the accuracy of potency determinations were investigated by repeating factorial 2×2 assays, in which one of the 2 sera being compared was a known dilution of the other. With the calculated regression lines not departing significantly from parallelism in each of 3 independent assays, the estimated potency ratios in 2 assays agreed with each other and with the true ratio. The remaining assay gave a potency ratio which significantly differed from the true value and was inconsistent with the other two estimates, as measured by chi-square. Taking this heterogeneity into account, the results of the 3 independent estimates were combined, and the average potency ratio did not significantly differ from the true ratio.

The applicability of the method to the comparison of different syphilitic sera was further examined in non-replicated tests. Regression lines were calculated for the data corresponding to 8 sera tested with 3 doses each, and a combined analysis of variance showed that the slopes of those lines did not depart from parallelism more than could be expected by chance alone.

The foregoing results indicate the possibility of assaying the antibody of syphilitic sera by complement fixation in a system of parallel straight lines.

359 J. O. IRWIN. The Study of the Physiological Effects of Hot Climates.

One of the physiological requirements for health is the maintenance of a practically constant body temperature. The environmental factors which affect the rate of heat loss are the temperature, humidity and rate of movement of the air and the radiation from the surroundings; but the rate of loss is largely governed by the physiological mechanisms which serve the body as thermostatic controls.

A single index of thermal environment known as Effective Temperature was designed to take account of the temperature, humidity, and rate of movement of the air and was a measure of subjective feelings of comfort. On account of defects in Effective Temperature McArdle and colleagues in London were led to construct an alternative index based on sweating rates. Using the results of nearly 1000 individual experiments they constructed an empirical nomogram from which the Predicted 4 hour Sweating Rate—*P4SR*—for any set of working conditions could be ascertained provided the environmental factors, the metabolic cost of the work and the clothing worn were known. It was desired to estimate the accuracy of the *P4SR* scale, and to assess the value of the Effective Temperature scale for grading the severity of thermal conditions in relation to human activities in the Tropics.

An experiment was carried out at Singapore to determine the effects on men naturally acclimatised to the Tropics of exposure for four hours twice weekly to varying combinations of air temperature, humidity and air movement. Combinations of air velocity, dry bulb and wet bulb temperature were designed to cover the same range as had been investigated in London. Originally a $4 \times 3 \times 2$ factorial arrangement had been suggested, but this was modified for technical reasons. There were 3 teams with 4 subjects in each—young naval ratings who volunteered from ships or shore establishments on the Far East Station. Each team had two 4-hour periods a week in the hot room. Four work-clothing combinations were tested at each exposure: Working in shorts, Working in overalls, Resting in shorts, Resting in overalls. Work consisted in step-climbing according to a certain routine. These four categories have been called "Postures" for convenience. They may be allocated to 4 subjects in 24 different ways, and one of these was assigned randomly to each of the 24 climate combinations, separate randomisations being used for each team.

In this plan all separate climate and posture comparisons were unconfounded with differences between persons and were therefore

equivalent to comparisons on the same persons. The error term for these was originally intended to be based on such climate-posture interactions as were not confounded with personal differences. It was not realised that these interactions were as important as they proved to be. Further, by what was subsequently recognized to be an error of judgment, the 12 subjects were, on the basis of a uniformity trial carried out before the main trial started, divided into four grades of sweating with three subjects in each and one member of each grade was put in each time. The arrangement of the climatic variables was thus not factorial, and it was not possible to allow for personal differences by the analysis of variance itself. The only way to correct for these was by analysis of covariance on the basis of the uniformity trial. To meet the other difficulty, the results of the trial were divided into two distinct sections, the first containing all combinations of 90° and 120° dry bulb temperature, 80° and 85° wet bulb temperature and the four air velocities, and the second all combinations of 90, 100, 120°F. dry bulb temperature and 80, 83, 85, 88°F. wet bulb temperature at the third air velocity (300 ft./min.). The analysis of covariance was carried out separately for the two sections. The combination of 90° and 120°F. dry bulb temperature with 80° and 85° wet bulb temperature at an air velocity of 300 ft./min. occurred in each.

The statistical analysis was carried out for a number of response variables—Total sweat loss, Total sweat loss per square metre of body surface, Evaporative water loss (absolute and per square metre), Final rectal temperatures, Final pulse rates (seated and standing), Comfort ratings and Efficiency ratings. Examples given refer to the variate "Total sweat loss".

Regression analysis was used to compare the "Total sweat rates" obtained from this series of experiments with the $P4SR$ values obtained from the nomogram constructed by McArdle and his colleagues and with Effective Temperature. If each work-clothing combination is taken separately, the predictive accuracies for these "naturally acclimatised" naval ratings of the Effective Temperature scales and the $P4SR$ nomogram are about the same, though there is a slight advantage to the latter in predicting sweat loss. However, when the results of all groups of experiments are combined, correlations with effective temperature are considerably lower than with $P4SR$, because the Effective Temperature scales make no allowance for differences in work rates. In this sense $P4SR$ is a more comprehensive index. It also gives a more adequate picture of the change in stress with air movement. On the other hand, the predicted 4 hour sweat rate can only be applied within the range of climate-work-clothing combinations which cause

people to sweat. It cannot replace Effective Temperature under the more comfortable and desirable conditions of light and sedentary work with which it was designed to deal primarily, for sweating will not occur under these conditions. In this sense it is less comprehensive than Effective Temperature but it is a more accurate index of physiological effect under conditions of thermal stress.

360 J. N. MANCEAU. **Application of the Covariance Analysis to the Comparative Study of Two Anthelmintics.**

Prevention and treatment of the several helminth infestations is one of the major concerns of public health officers in the Amazon region. A test was made to compare the efficiency of Aralen a new drug, with that of Hexylresorcinol, the one commonly used in the treatment of infected persons.

A sample of 74 children was chosen at random from the Lauro Sodre professional school (Belem, State of Para), and a stool specimen was taken from each child to determine the degree of infestation (number of eggs per centigram of feces) with A. lumbricoides, Ancylostoma, and T. trichiura, before treatment.

The sample of 74 children was arranged into 37 pairs, each pair having, insofar as possible, the same degree of infestation by A. lumbricoides. One child selected at random from each pair was treated by Hexylresorcinol, and the other was treated by Aralen.

The results were subjected to the analysis of covariance. Hexylresorcinol proved to be better than Aralen in the treatment of infection with A. lumbricoides and Ancylostoma. No significant difference was observed in the treatment of infection with trichiura.

The independent variable used (degree of infestation before treatment) made it possible to attain a higher degree of accuracy in the experiment.

361 A. E. BRANDT AND GILBERT H. FLETCHER. (Biometrician, Health and Safety Laboratory, New York Operations Office, U. S. Atomic Energy Commission, and M.D., Pathologist, M.D. Anderson Hospital, The University of Texas, Houston, Texas. **Design of a Clinical Investigation of Very High Voltage Sources in the Radiotherapy of Cancer.**

The full title of this contribution should read, the design of a clinical investigation of the differences in reaction and clinical response between Cobalt-60 (1.2 Mev) therapy and 22 Mev Betatron therapy in the

treatment of cancers which are infrequently curable by conventional radiotherapy techniques. The two phases of this problem (the medical and the biometric) are defined and the paramount importance of the medical phase due to the use of human subjects is pointed out. The responsibility of the medical leader of this investigation for the medical excellence of the design as well as for the diagnosis and treatment of patients is presented. The objectives of the biometrician and the controls by which he achieves these objectives within the medical framework provided by the medical leader are given. The biometric design of that portion of the investigation relating to cancer of the cervix is presented as an example.

362 A. CHARBONNIER, B. CYFFERS, D. SCHWARTZ, A. VESSEREAU. **Application of Discriminatory Analysis to Medical Diagnostic.**

On individuals which belong to one among two or several families, measurements of several characters, or variables, have been made. The point is to find the linear functions of these variables—namely the "discriminant functions" by which we can allot each individual to its proper family with a minimum of risk of error.

Hypotheses are that, within each family, the distribution of the variables is a normal distribution with p variables, and that all these distributions have the same dispersion-matrix, but differ only regarding their centres.

In the following application of the method to the medical diagnostic, the families differ by the nature of a basic disease named "ictere". The variables are different constituents of the blood serum:

$$\text{albumin} = A \qquad \text{globulin} = \alpha_1, \alpha_2, \beta, \gamma, \sigma, \epsilon$$

$$p = \text{total mass of proteins.} \quad p = p - \sigma - \epsilon.$$

The values of these elements, measured by the technics of electrophoresis were recorded on a group of 197 ill persons.

The study was made in several steps.

First step: Observation of evident differences between the families, concerning each variable. This empirical information can be used to predict the nature of the disease. But, in better circumstances, diagnoses were made only in 40% of cases with 8 errors (or 20% of the diagnoses).

Second step: by t test of Student-Fisher, verification of significant differences between the families, for each variable of the electrophoresis.

Third step: Choice of the best variables for the establishment of the discriminant functions. The choice was possible between x (raw variable), $\log x$, x/P' or x/P. The values of t (comparison between hepatites and cancers) show that it is better to work with the variables x/P'. With six variables, plus the total mass of protides P, the computation of the coefficients of discriminant functions would have been very tedious. Only 4 variables were retained, which, regarding the values of t, were a priori the best for discrimination.

Fourth step: Computation of the discriminant functions. With 3 families, one discriminant function only is available if the centres of the families are on the same straight line. That was not the case, so it was decided to make the discrimination in two steps:

first, discrimination between "medicaux" and "chirurgicaux" (calculs + cancers) which is the most important,

then, among the "chirurgicaux", discrimination between "calculs" and "cancers".

Computations have been made with the data of the 197 diseased. Statistical tests show that, for the first function, the coefficients are significant, or almost significant, and that for the second, the coefficients are not very significant.

In order to allot *each diseased* to one of two families, a critical value is chosen. If, for a particular diseased, the value of the discriminant function is smaller than the critical value, the conclusion is: "family 1"— in the other case the conclusion is "family 2". With the best critical value, it is found that the theoretical % of errors is very high: 35%. In fact on the 197 diseased, 62 errors (or 32%) were found.

It is better to fix a priori the probability of error. With this position a segment (ab) is determined, and between the values a and b, no diagnosis is pronounced.

5% was chosen as probability of error for each of the discriminant functions separately—When the two functions are successively applied, with the possible responses:

No diagnosis
Hepatitis
Calcul or cancer
Calcul
Cancer

the probability of error is theoretically higher. The theoretical proportion of diagnoses (considering calcul or cancer as a diagnostic) is 45%; on the 197 diseased we find 41%.

Applying successively the two functions to the 197 diseased, we obtain the following results:

81 diagnoses (42%),
10 errors (5% of the total, 12% of the diagnoses).

There is a very good concordance between theoretical and observed values.

But a better confirmation of the validity of the method is the following. The discriminant functions were applied to another group of 81 diseased and the classification operated by the functions was entirely in harmony with the theory and with the results of the first group of diseased.

In conclusion, we can emphasize the following points:

1°)—This method is better than the empirical method: more responses, less errors. Its application is easy: tables which have been constructed, give very easily the values of the discriminant functions.

2°)—The method requires some precautions: it depends on the technics applied, and perhaps on the origin of the diseased.

3°)—The method has been good for the discrimination between "icteres médicaux" and "chirurgicaux", but fairly poor for the discrimination between "calculs" and "cancers". An improvement of the method can be expected, taking account of the sex, the age, of the ill-persons.

4°)—It seems that the same method would be of interest in many other cases of medical diagnostic.

363 ARTHUR LINDER. **On a Particular Kind of Grazing Experiment.**

Research has been carried out on the effect of fertilizers on pastures in the higher regions of the Grisons (Switzerland). Grazing experiments were set up to evaluate the palatability of grass on fertilized plots. One typical experiment consisted of five blocks with six plots receiving different kinds and amounts of fertilizers. A fence was drawn around the area and two cows were allowed to graze for two hours. The effective grazing times were recorded. Analysis of the results shows significantly longer grazing on plots with higher levels of fertilizers. The experiment was repeated after one year without changing the treatment of plots. Results of the two trials agreed closely.

THE BIOMETRIC SOCIETY

Italian Region. On April 22, 1955, the Italian Region held its fifth annual meeting in Pavia in a joint session with the Italian Genetic Society (Associazione Genetica Italiana). Three papers were presented. The first, by F. Brambilla and L. L. Cavalli-Sforza, on Biological variability of environmental origin, concerned a model for the non-genetic variability in a population and its statistical consequences. The method is based upon a transformation of the frequency distribution of the environmental stimuli by a function connecting the stimulus and the intensity of biological response. When this function has a maximum or minimum, the frequency distribution of the response shows some peculiar characteristics which were discussed. The second paper by A. Previtera concerned the measurement of frailty in children. Two established auxological indexes were examined statistically with a population of children of various ages and an improved index suggested. The third paper by E. Baldacci, G. Fogliani and E. Betto described the planning of experiments on the control of *Peronospora* by fungicides. The problem here lay in the fact that artificial infection with this species is not easy, so that field experiments have to be based upon natural epidemics.

British Region. The Region held its twenty-fourth meeting at the Wellcome Research Institute in London on May 4, 1955. In the first contribution, J. A. Fraser Roberts discussed the supposed difference between the sexes in variability of intelligence. That boys are more variable than girls in their scores on intelligence scales seems to arise empirically from the data. After excluding a large volume of unsuitable records, this greater variability of boys seems to be real, at least in some kinds of tests and at certain ages. The difference may be a biological phenomenon, depend on the way test scales are constructed, or be due to differences between boys and girls in education and in their social and home environments. The problem is being examined by analyzing a number of different samples. Dr. Roberts' paper was followed by a discussion opened by J. W. Craven and P. Olden on some difficulties in sequential analysis.

Switzerland. The Swiss Section of the Biometric Society met jointly with the Swiss Society of Genetics at the University of Berne on May 22, 1955. The following papers were presented: S. Rosin, Statistical problems in the evaluation of blood-group determinations; H. L. LeRoy, Mathematical statistics as an aid in the solution of problems of selection in animals; A. Kaelin, Influence of selection upon

the estimation of gene frequencies among siblings in human genetics; and A. Linder and B. Grab, A statistical study on the relation between several anthropological measures in the infant and the corresponding measurement in its parents.

Region Francaise. Lors de la reunion de la Societe Francaise de Biometrie, qui eut lieu mardi le 24 Mai a l'Ecole Normale Superieure a Paris, l'ordre du jour etait le suivant: R. Turpin et M. P. Schutzenberger, Remarques sur un probleme de consanguinite; et Dr. Charbonnier, B. Cyffers, D. Schwartz et A. Vessereau, Discrimination entre icteres medicaux et chirurgicaux a partir des resultats de l'analyse electrophoretique des proteines du serum.

WNAR. On the first day of the annual meeting of the Wester North American Region in Pasadena, California, on June 23, 1955, the program consisted of two scientific sessions and a luncheon business meeting. The morning session on Ecology, under the chairmanship of R. O. Erickson, offered the following papers: P. E. Fields, Factorial designs and the guidance of downstream migrant salmon and steelhead trout; D. G. Chapman and R. Pyke, Statistical theory of some migration population models; J. L. Baily, Jr., Variation of the *Pectin gibbus* complex; and R. F. Tate and R. L. Goen, Minimum variance unbiassed estimation for a truncated Poisson parameter. During the luncheon business meeting, following other business, the Region voted to elect future officers by mail ballot. At the afternoon session on Psychometrics, J. P. Guilford presided and the following papers were presented: H. D. Kimmel, Reliability of qualitative, categorical judgements; D. A. Grant, Statistical tests in the comparison of curves; P. R. Merrifield, Quantification of ordering behavior; J. A. Gengerelli, Methods of constellation analysis; H. H. Harman, Some observations on factor analysis; and J. W. Frick, Effect of varied interpolated stimuli upon the time-order error. At the closing section on Genetics on June 24, A. H. Sturtevant was in the chair and the following papers were read: L. A. Lider, A group of long-term, perennial and non-replicated rootstock trials; C. N. Stormont, Estimates of frequencies of B-alleles in three breeds of dairy cattle; G. E. Dickerson, Some unsolved statistical problems of importance in quantitative genetics; and H. Rubin, Axiomatization of genetics.

Japan. Members of the Biometric Society in Japan held their second spring meeting in Tokyo on April 5, 1955. Five papers were presented: M. Masuyama, Microbiological inspection of bulk material by composite sample; M. Kiyoku, On the response-surface in the interaction of temperature and dryness upon insects; T. Okuno and T. Sasaki, Factorial analysis of the adjusted treatment means obtained

by analysis of covariance; T. Yamada, On the chance distribution of quantitative characters in a population due to interplant competition; and S. Hashiguchi, Estimating the variance component for the evaluation of heritability. To insure their wider distribution, the Chapter has printed an extended summary of each paper in a 34-page Japanese booklet, in which three of the papers have been provided with English abstracts. The preceding spring session of the Japanese Chapter met jointly with the Japanese Agricultural Society.

Australasian Region. The Region held its fifth meeting at the University of Melbourne on August 22, 1955, as part of the 31st meeting of the Australian and New Zealand Association for the Advancement of Science. The following program was cosponsored by the Zoology Section of the Association: R. T. Leslie, A statistical approach to the physiological problem of thresholds; E. J. Williams, Sidelights of sampling surveys; G. S. Watson, Missing and mixed-up values in contingency tables; P. J. Claringbold, Discriminant analysis in the interpretation of semi-quantal data; and (Mrs.) G. L. Richardson and F. E. Binet, Discriminant analysis of species of the genus Murravia (Brachiopoda, Tertiary, Recent).

ENAR. The Eastern North American Region met in East Lansing, Michigan, on September 6–8, 1955, as part of the annual meeting of the American Institute of Biological Sciences. The program was arranged by a committee consisting of E. L. Green (Chairman), M. Whittinghill, T. Park and W. C. Jacob. The Local Representative was W. D. Baten.

A joint session with the Genetics Society of America and the American Society of Human Genetics on the afternoon of September 6 was titled, "Sewall Wright's Contributions to Population Genetics". Professor Wright was present and received a standing ovation by the several hundreds in attendance. The following papers were presented under the chairmanship of Paul R. David: C. C. Li, The concept of path coefficients and their impact on population genetics; J. F. Crow, Effective population number, its estimation and relevance as one factor in evolution; H. B. Glass, Some evidence for genetic drift in human populations; W. P. Spencer, Natural populations of Drosophila and the Wrightian model of evolution; and J. P. Scott, The analysis of quantitative traits in populations.

The following afternoon a joint session with the American Society of Horticultural Science was devoted to "Sampling Applications in Horticulture" with J. C. Jacob in the chair and the following papers: J. P. McCollum, Sampling tomato fruits for composition studies; W. C. Kelly, Sampling vegetative portions of vegetable plants for vitamin

analysis; W. W. Jones, Sampling citrus and avocado trees for nutritional studies and yield relationships; N. J. Shaulis, Sampling small fruit for composition and nutritional studies; and J. A. Rigney, Sampling soil for composition studies.   A Biometric Society Dinner in the evening was well attended.

The closing session on September 8, held jointly with the Ecological Society of America and the American Society of Naturalists, concerned "Quantification in Population Ecology", with Thomas Park as chairman. The program consisted of the following papers: J. Neyman, Statistical models of population phenomena; D. E. Wohlschlag, Conceptual problems in the application of theoretical models to unstable fish population; and L. C. Cole, Inductive procedures in quantitative ecology.

Finances for 1954. Following the practice initiated last year, the following audited financial statements for BIOMETRICS and for the office of the Secretary-Treasurer for 1954, are listed under headings similar to those used last year.

<div align="center">BIOMETRICS</div>

Income

Reserve from sale of No. 1 in Vols. 3 and 7	$ 1321.50	
Balance from 1953 books	4993.95	$ 6315.45

Member subscriptions

ASA, 394	1576.00	
Biometric Society, 1144	3936.00	5512.00

Non-member subscriptions, 622		4352.25

Sale of back issues

Vol. 1–5, ,864.40; Vol. 6–10, ,1195.30	2059.70	
Vol. 3, No. 1, ,217.50; Vol. 7, No. 1, ,66.00	283.50	2343.20

Sale of reprints		1124.61
Insurance overpayment (refund)		26.56
Total Income		$19674.07

Expenses

Overpayments and cancellations	$	47.50
Supplies, miscellaneous		38.27
Farm Bureau, insurance on back issues		48.96
Postage		300.00
Institute of Statistics, Editorial Management		1000.00
ASA, ½ net profit Vols. 1–5 (5th and final payment)		432.20

Wm. Byrd Press

Production of Biometrics, Vol. 10, 1954	$ 8582.00	
Offprints of Dec. 1953 to Sept. 1954	1232.80	9814.80

Total Expenses		$11681.73
Balance		$ 7992.34

Office of the Secretary-Treasurer

Income

Subscriptions, 1953—$369.50, 1954—$2913.75	$ 3283.25
Dues, 1953—$97.38, 1954—$1993.75	2091.13
Sustaining memberships, 1954	500.00
Back dues and subscriptions	41.50
Reprints, directories, back issues	14.45
Sale of excess equipment, bank charges	100.19
Overpayments ($122.16 less 55.91 for credits taken)	66.25
Total Income	$ 6096.77

Expenses

BIOMETRICS, 1954—$2262.50, 1953 and earlier—$49.82	$ 2312.32
Regional allotments	340.75
Salaries and special services	1438.52
Printing, stationary, supplies, etc.	159.56
Postage	136.48
Telephone	86.82
Depreciation	230.89
Total Expenses	$ 4705.34
Excess of Income over Expense	$ 1391.43

Balance Sheet as of December 31, 1954

Assets

Cash on hand (including petty cash) , ...	$ 2908.73

Liabilities

Dues and Subscriptions for 1955 . . . . $	218.50	
Surplus, Jan. 1, 1954 (excluding funds in transit)	1298.80	
Gain for period*	1391.43	$ 2908.73

*Owed against this gain but not paid in 1954—BIOMETRICS $1822.07, regional allotments $2.50; owing to this office for 1954 from Regional Treasurers and National Secretaries but not received in 1954—subscriptions $947.00, dues $261.75.

BIOMETRICS

Contents of Volume XI